Cambridge studies in advanced mathematics 186

The Bellman Function Technique in Harmonic Analysis

VASILY VASYUNIN
ALEXANDER VOLBERG

The Bellman Function Technique in Harmonic Analysis

The Bellman function, a powerful tool originating in control theory, can be used successfully in a large class of difficult harmonic analysis problems and has produced some notable results over the last 30 years. This book by two leading experts is the first devoted to the Bellman function method and its applications to various topics in probability and harmonic analysis. Beginning with basic concepts, the theory is introduced step-by-step starting with many examples of gradually increasing sophistication, culminating with Calderón–Zygmund operators and endpoint estimates. All necessary techniques are explained in generality, making this book accessible to readers without specialized training in nonlinear PDEs or stochastic optimal control. Graduate students and researchers in harmonic analysis, PDEs, functional analysis, and probability will find this to be an incisive reference, and can use it as the basis of a graduate course.

Vasily Vasyunin is Leading Researcher at the St. Petersburg Department of Steklov Mathematical Institute of Russian Academy of Sciences and Professor at the Saint Petersburg State University. His research interests include linear and complex analysis, operator models, and harmonic analysis. Vasyunin has taught at universities in Europe and the United States. He has authored or coauthored over 60 articles.

Alexander Volberg is Distinguished Professor of Mathematics at Michigan State University. He was the recipient of the Onsager Medal as well as the Salem Prize, awarded to a young researcher in the field of analysis. Along with teaching at institutions in Paris and Edinburgh, Volberg also served as a Humboldt senior researcher, Clay senior researcher, and a Simons fellow. He has coauthored 179 papers, and is the author of *Calderón–Zygmund Capacities and Operators on Non-Homogenous Spaces*.

The Bellman Function Technique in Harmonic Analysis

VASILY VASYUNIN
Russian Academy of Sciences

ALEXANDER VOLBERG
Michigan State University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108486897
DOI: 10.1017/9781108764469

First published 2020

Printed in the United Kingdom by TJ International, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Vasyunin, Vasily I., 1948– author. | Volberg, Alexander, 1956– author.
Title: The Bellman function technique in harmonic analysis /
Vasily Vasyunin, Alexander Volberg.
Description: Cambridge ; New York, NY : Cambridge University Press, 2020. |
Includes bibliographical references and index.
Identifiers: LCCN 2019042603 (print) | LCCN 2019042604 (ebook) |
ISBN 9781108486897 (hardback) | ISBN 9781108764469 (epub)
Subjects: LCSH: Harmonic analysis. | Functional analysis. | Control theory.
Classification: LCC QA403 .V37 2020 (print) | LCC QA403 (ebook) |
DDC 515/.2433–dc23
LC record available at https://lccn.loc.gov/2019042603
LC ebook record available at https://lccn.loc.gov/2019042604

ISBN 978-1-108-48689-7 Hardback

To our parents

Contents

	Introduction	*page* x
I.1	Preface	x
I.2	Acknowledgments	xii
I.3	The Short History of the Bellman Function	xii
I.4	The Plan of the Book	xv
I.5	Notation	xvi
1	**Examples of Exact Bellman Functions**	1
1.1	A Toy Problem	1
1.2	Buckley Inequality	7
1.3	John–Nirenberg Inequality	14
1.4	Homogeneous Monge–Ampère Equation	28
1.5	Bellman Function for General Integral Functionals on BMO	30
1.6	Dyadic Maximal Operator	43
1.7	Weak Estimate of the Martingale Transform	62
1.8	Burkholder's Bellman Function	80
1.9	On the Bellman Function Generated by a Weak Type Estimate of a Square Function	113
1.10	More about Buckley's Inequality	137
1.11	Hints and Answers	144
2	**What You Always Wanted to Know about Stochastic Optimal Control, but Were Afraid to Ask**	150
2.1	Disclaimer	150
2.2	Stochastic Integrals Are Not That Simple	150
2.3	Itô's Definition of Stochastic Integral	153
2.4	Stochastic Differential and Itô's Formula	155

2.5 Bellman Functions of Stochastic Optimal Control Problems and Bellman PDEs 156
2.6 Almost Perfect Analogy between Stochastic Optimal Control and Harmonic Analysis: Functions of One Variable 164
2.7 Almost Perfect Analogy between Stochastic Optimal Control and Harmonic Analysis: Functions on the Complex Plane 182
2.8 A Problem of Gohberg–Krupnik from the Point of View of Stochastic Optimal Control 195

3 Conformal Martingale Models: Stochastic and Classical Ahlfors–Beurling Operators 206
3.1 Estimates of Subordinated Martingales 206
3.2 Conformal Martingales and the Ahlfors–Beurling Transform 208
3.3 Proof of Theorem 3.2.3: Right-Hand Side Conformality, $2 < p < \infty$ 210
3.4 Proof of Theorem 3.2.3: Left-Hand Side Conformality, $1 < p < 2$ 226
3.5 Burkholder, Bellman, and Ahlfors–Beurling Operator in L^p for Large p 232

4 Dyadic Models: Application of Bellman Technique to Upper Estimates of Singular Integrals 247
4.1 Dyadic Shifts and Calderón–Zygmund Operators 247
4.2 Sharp Weighted Estimate of Dyadic Shifts 266
4.3 Universal Sufficient Condition: Boundedness of All Calderón–Zygmund Operators in Two Different Weighted Spaces 292

5 Application of Bellman Technique to the Endpoint Estimates of Singular Integrals 334
5.1 Endpoint Estimate 334
5.2 The Bellman Function of Weak Weighted Estimate of the Martingale Transform and Its Logarithmic Blow-Up 337
5.3 Sharp Weak Weighted Estimate for the Martingale Transform 353
5.4 Obstacle Problems for Unweighted Square Function Operator: Burkholder–Gundy–Davis Function 367
5.5 Bollobás Function 383
5.6 The Weak Norm of the Square Function 395
5.7 Saturation of Estimates by Extremal Sequences 398
5.8 An Obstacle Problem Associated with the Chang–Wilson–Wolff Theorem 400
5.9 Strong Weighted Estimate of the Square Function 402

5.10 Weak Weighted Estimates for the Square Function 411
5.11 Restricted Weak Weighted Estimate of the Square Function 432

References 434
Index 443

Introduction

I.1 Preface

The subject of this book is the use of the Bellman function technique in harmonic analysis. The Bellman function, in principle, is the creature of another area of mathematics: control theory. We wish to show that it can be used very successfully in a big class of harmonic analysis problems. In the last 25–30 years some outstanding problems in harmonic analysis were solved by this approach. Later, 10–15 years after, another solution has been found by more classical methods involving some highly nontrivial stopping time argument.

This is what happened with the A_2 conjecture and then very recently with the A_1 conjecture concerning weighted estimates of singular integrals. Some other problems solved by the Bellman function method still await their "de-Bellmanisation." Among such problems, we can list the celebrated solution by Burkholder of Pełczyński's problem about Haar basis, the best L^p estimates of the Ahlfors–Beurling operator, and many matrix weight estimates.

One of the main technical advantages of the Bellman function technique is that it does not require the invention of any sophisticated stopping time argument of the kind that is so pervasive in modern harmonic analysis. We can express this feature by saying that the Bellman function knows how to stop the time correctly, but it does not show us its secret.

The purpose of this book is to present a wide range of problems in harmonic analysis having the same underlying structure that allows us to look at them as problems of stochastic optimal control and, consequently, to treat them by the methods originated from this part of control theory.

We intend to show that a certain class of harmonic analysis problems can be reduced (often without any loss of information) to solving a special partial differential equation called the Bellman equation of a problem. For that purpose, we first cast a corresponding harmonic analysis problem as a stochastic optimization problem.

A quintessentially typical problem of harmonic analysis is to find (or estimate) the norm of this or that (singular) operator in function space L^p. If we think about the operator as a black box, we should think about the unit ball of L^p as its input. The unit ball of L^p is not compact in norm topology, so there is a priori no extremizer, and moreover, it seems to be very difficult to "list" all functions in the unit ball and "try" them as black box inputs one by one.

The stochastic point of view helps here, because we can think about input as a stochastic process stopped at a certain time. This point of view gives a very nice and powerful way to list all inputs as solutions of simple stochastic differential equations with some unknown stochastic control. Then the norm of the operator becomes a functional on solutions of stochastic differential equations that we need to optimize by choosing optimal control.

The technique of doing that is to consider the Bellman function of this control problem and to write the Hamilton–Jacobi–Bellman equation whose solution the Bellman function is supposed to be.

This book can be used as the basis of a graduate course, and it can also serve as a reference on many (but not all) applications of the Bellman function technique in harmonic analysis.

A certain number of very important results obtained with the use of the Bellman function technique stayed outside of the scope of this book. For example, these are the twisted paraproducts results of V. Kovac [**91, 92**], and, in general, the applications of the Bellman function to multilinear and nonlinear harmonic analysis. In the last category one finds the works of C. Muscalu, T. Tao, and C. Thiele [**119**] concerning nonlinear analogs of the Hausdorff–Young inequality that relates the norm of a function and the norm of its nonlinear Fourier transform. This book does not present the recent results of O. Dragicevic and A. Carbonaro [**33, 34**], where the authors study universal multiplier theorems in the setting of symmetric contraction semigroups. In particular, the authors solved a long-standing problem of finding the optimal sector, where generators of symmetric contraction semigroups always admit a H^∞-type holomorphic functional calculus on L^p. This is done by a subtle application of the Bellman function technique on a flow that is given by the semigroup. Numerous multiplier theorems are improved due to this result, and new results on pointwise convergence related to a symmetric contraction semigroup on a closed sector are obtained.

The corresponding papers can be found in the References, and the reader is encouraged to study these beautiful applications of the Bellman function ideology.

I.2 Acknowledgments

Our foremost acknowledgments are to our teachers M. S. Birman and N. K. Nikolski. We also learned most of what we know from our collaborators, colleagues, and our students: Fedor Nazarov, Sergei Treil, Igor Verbitskiy, Dmitriy Stolyarov, Pavel Zatitskiy, Stefanie Petermichl, Paata Ivanisvili, Alexander Reznikov, Oliver Dragicevic, Irina Holmes, Leonid Slavin, and many others. The ideas we caught while working and/or talking with them are incorporated in this book. Special thanks are due to Marina Kuznetsov who corrected our most horrible blunders in English.

We are very grateful to Oberwolfach Mathematics Research Institute, to Centre International de Rencontres Mathématiques in Luminy, and to Department of Mathematics of Michigan State University for excellent working conditions and invitations for the two of us together to stay during 2014–2017. We are also grateful to NSF support that allowed us to work together and meet regularly.

Finally, we would like to thank our friends and relatives who encouraged us during writing of this book, especially our wives, Nina Vasyunina and Olga Volberg.

I.3 The Short History of the Bellman Function

The Bellman equation and the Euler–Lagrange equation both deal with extremal values of functionals. Given a functional, the Euler–Lagrange approach gives us a differential equation that rules the behavior of extremizers of a functional in question. The Bellman equation has quite a different nature. At first glance, it does not give any information on the extremizers. Moreover, typically, there will be no extremal function. Only a sequence of almost extremizers typically exists, which is yet another difference with applications of the Euler–Lagrange equation.

In the Bellman paradigm, the (system of) differential equation(s) is given to us, and we do not need to find it as in the Euler–Lagrange approach. However, the given differential equation (or a system of equations) also has an unknown functional parameter called control. We have to find the best control in the sense that this control will optimize a functional applied to the solution of (an) already given (system of) differential equation(s).

So the idea of Bellman is amazingly striking and incredibly simple simultaneously. Let us denote by $\mathbf{B}(x)$ the extremal value of the functional that we want to optimize on solutions of a given system of differential equations with initial value at x at initial time 0. Using the fact that at time Δt, we "know" where the solution is, and that having started at $x(\Delta t)$ this solution is also extremal (if the control is chosen correctly), one can deduce a partial differential equation on $\mathbf{B}$.

This equation is called the Hamilton–Jacobi equation. Due to the functional in need of optimization, and depending on the system of differential equations that is given to us, the Hamilton–Jacobi equation varies, but it stays in a certain class of first-order nonlinear (usually) PDEs.

However, if the system of differential equations mentioned previously is not the usual system, but is a system of stochastic differential equations, and we are asked to optimize not just a functional of solutions of this system, but the expectation (the average) of this functional, then the scheme mentioned previously can be applied as well. The resulting PDE is often called the Hamilton–Jacobi–Bellman equation, and the presence of stochasticity makes it a second-order nonlinear (usually) PDE. It belongs to the class of equations called degenerate elliptic equations, see, e. g., [**120**].

Previously we presented a very short exposition of the Bellman function of stochastic optimal control. In this branch of mathematics, it is also often called the value function. The reader who wants to acquire good knowledge in this area is advised to read [**95**].

But the goal of the book is to show the deep (and almost perfect) analogy between the Bellman function technique in stochastic optimal control and the Bellman function technique in harmonic analysis, which is the branch of analysis dealing with the estimates of singular integrals.

It was arguably in [**130**], and especially in [**131, 193**], where this parallelism between a wide class of harmonic analysis problems and the stochastic optimal control got recognized at face value.

The observation of this parallelism between two different branches of mathematics is sort of important for this book, but the ideas that now are generally recognized as the "Bellman function technique" in harmonic analysis have been around long before those papers.

Without any claim of completeness, we can list several articles and their ideas that now can be recognized as instances of application of the Bellman function technique (without ever mentioning stochastic control, the value function, or anything like that).

In the area of probability theory that deals with optimal problems for Brownian motion or for martingales, D. Burkholder [**22–31**], B. Davis [**50**], and Burkholder–Gundy [**32**] used what we call now the Bellman function as their main tool of finding the constants of best behavior of stopping times and of martingales with various restrictions.

But to the best of our knowledge, the first use of the idea underlying the Bellman function technique is due to A. Beurling, who found the exact function of uniform convexity for the space $L^p(0,1)$. Strangely enough, his work was not published: Beurling just made an oral report in Uppsala in 1945, and the exposition of his idea can be found in the paper of O. Hanner [**66**] of 1956. However, Beurling used certain magic guesses. These guesses were explained in the paper of Ivanisvili–Stolyarov–Zatitskiy [**82**], who showed that Beurling's function method is nothing other than perhaps the first occasion of the application of the Bellman function technique in harmonic analysis.

We think that chronologically the next case of using the Bellman function technique in harmonic analysis was again related to uniform convexity. But this case deals with the general theory of Banach spaces. In 1972, P. Enflo in [**59**] proved that the Banach space X is super-reflexive if and only if it can be given an equivalent norm that is uniformly convex. In fact, the "if" part was proved by R. C. James in [**86**]. Enflo proved the "only if" part, and the proof of Lemma 2 of [**59**] now reads as a typical Bellman function technique proof.

In 1975, G. Pisier [**154**] gave another proof of the James–Enflo result that the Banach space X is super-reflexive if and only if it has an equivalent uniformly convex norm. His proof used X-valued martingale interpretation of super-reflexivity. The uniformly convex norm on X was constructed in the second line of the proof of Theorem 3.1 of [**154**]. In fact, for a vector $x \in X$ this equivalent norm $|x|$ is defined as the infimum of a certain functional on X-valued martingales starting at x, and this is a quintessential Bellman function definition.

Let us briefly explain why we associate such an approach (also used in all the papers of Burkholder, Gundy, Davis mentioned previously) with the Bellman function technique described previously.

Roughly speaking, any martingale is a solution of a controlled stochastic differential equation (with continuous or discrete time), where martingale differences play the role of control that should be optimized to give a prescribed functional on martingale the "best" value. In the case explained in [**154**], the functional is given at the beginning of the proof of Theorem 3.1, and its optimal value is precisely $|x|$ – the equivalent norm of the initial vector x, where a martingale (the solution of a stochastic differential equation in our interpretation) has started.

The fact that $x \to |x|$ is uniformly convex is exactly the Hamilton–Jacobi–Bellman PDE, as the reader will conclude after reading this book. It sounds strange: Why should a certain inequality be called a partial differential equation?

It will be explained repeatedly that the Bellman PDE pertinent to a harmonic analysis problem is quite often, in fact, a certain second-order finite difference inequality.

It is difficult to find the optimal solutions of inequalities, so the reader will see in a case-by-case study how we account for this difficulty and how we remedy it.

In probability theory, there was an interest in understanding the relationship between the various norms of the stopping time T and the corresponding norms of $W(T)$, where W is the Brownian motion. The exposition of these results (and related martingale results) of B. Davis [**50**], G. Wang [**196**], and [**197**] can be found in Chapter 5.

A huge amount of work has been done in the papers by D. Burkholder [**21, 22**] and by Burkholder and Gundy [**32**]. These are all Bellman function technique papers. In particular, this method (without mentioning any stochastic optimal control or Hamilton–Jacobi equation) was used by Burkholder (in his seminal articles cited previously) to solve problems of A. Pełczyński, concerning sharp constants for unconditional Haar basis in L^p. We adapt Burkholder's solution to our language of the Bellman function technique, which is done in Section 1.8. This is one of those cases when it is easy to write the Bellman equation but difficult to solve it.

I.4 The Plan of the Book

In Chapter 1, we give nine precise Bellman functions corresponding to several typical harmonic analysis problems. As we already mentioned, Section 1.8 of this chapter is devoted to Burkholder's Bellman function. The John–Nirenberg inequality presents a very nice model for the application of the Bellman function technique, which the reader will find in Section 1.3. Then, in Section 1.5 we extend the method of the Bellman function to rather general functionals on the space BMO.

In Chapter 2, we first list elements of stochastic calculus and introduce the Bellman function of stochastic optimal control. Then in Section 2.6, we collect examples that show the perfect analogy between stochastic optimal control and a wide class of harmonic analysis problems. After that, we turn our attention to a class of problems from complex analysis that also can be adapted to the Bellman method. One of these problems is finding Pichorides constants, yet another question is concerned with the solution of Gohberg–Krupnik problem by B. Hollenbeck and I. Verbitsky [**69**]. Our main goal is to show that all harmonic analysis problems in this chapter can be interpreted as problems of stochastic optimal control. An important disclaimer should be made: the stochastic optimal control point of view helps us to write down a correct Bellman partial differential equation, but it does not, in any sense, help to solve it. And solving it can be a major difficulty.

Chapter 3 is devoted to sharp estimates of conformal martingales. We then use these results to consider one particular singular integral, the Ahlfors–Beurling transform. We give the best up-to-date estimates of this transform.

Chapter 4 demonstrates an interesting and unexpected feature of the Bellman function technique. Namely, it has been noticed that the Bellman function built for one problem can be used in another problem, sometimes not too close to the original problem. This allows us to use the Bellman functions for the weighted martingale transform to have the right estimates for much more complicated dyadic singular operators, the so-called dyadic shifts. Moreover, one need not know the precise form of the Bellman function, one should just know of its existence. This idea for the Ahlfors–Beurling transform was used by S. Petermichl and A. Volberg in [**152**]. In that chapter, we follow the ideas of S. Treil [**181**] with a slight modification.

It has been noticed repeatedly that the Bellman function technique can be used not only to prove the conjectural estimates of singular integrals, but also to disprove the estimates. This is, roughly speaking, the consequence of the fact that the language of Bellman functions is often exactly adequate and equivalent to harmonic analysis problems for which these functions are built. This observation helps to find sharp constants in several endpoint estimates for singular integrals. That point of view also brings counterexamples to several well-known conjectures. We devote Section 5.2 to such counterexamples. The rest of this chapter is devoted to using the Bellman function technique to find sharp estimates in several classical problems concerning the square function operator. Even though the sharp constants for this operator have been studied since 1975, there are still open questions and we discuss them in Chapter 5.

I.5 Notation

We conclude this Introduction by a short list of notation that will be used throughout the whole book.

The average of a summable function w over an interval I will be denoted by the symbol $\langle w \rangle_I$:

$$\langle w \rangle_I \stackrel{\text{def}}{=} \frac{1}{|I|} \int_I w(t)\, dt\,,$$

where $|I|$ stands for the Lebesgue measure of I.

We introduce the Haar system, normalized in L^∞:

$$H_I(t) = \begin{cases} -1 & \text{if } t \in I_-, \\ \ \ 1 & \text{if } t \in I_+, \end{cases} \tag{0.1}$$

and another one, normalized in L^2:

$$h_I(t) = \frac{1}{\sqrt{|I|}} H_I(t).$$

Then

$$|I|(\langle w\rangle_{I_+} - \langle w\rangle_{I_-}) = 2(w, H_I)$$

and

$$\sqrt{|I|}(\langle w\rangle_{I_+} - \langle w\rangle_{I_-}) = 2(w, h_I).$$

The characteristic function of a measurable set E is denoted by $\mathbf{1}_E$.

The symbol $\mathcal{D}$ stands for a dyadic lattice, and $\mathcal{D}_n$ stands for the grid of intervals (or cubes) of length (or side-length) 2^{-n}, $n \in \mathbb{Z}$. The σ-algebra generated by $\mathcal{D}_n$ is denoted by $\mathcal{F}_n$.

The symbol $\mathbb{E}_n$ stands for the expectation with respect to the σ-algebra $\mathcal{F}_n$. Then Δ_n stands for $\mathbb{E}_{n+1} - \mathbb{E}_n$, $\Delta_I \stackrel{\text{def}}{=} \mathbf{1}_I \Delta_n$ for $I \in \mathcal{D}_n$, and thus,

$$\Delta_n = \sum_{I \in \mathcal{D}_n} \Delta_I \,.$$

Bellman functions are usually denoted by $\mathbf{B}$, but in Sections 5.4–5.7 they are denoted by $\mathbf{U}$ to follow the established tradition coming from probability.

The matrix of second derivatives of a function B on $\mathbb{R}^d$ (the Hessian matrix) is denoted by $\frac{d^2 B}{dx^2}$ or H_B. The symbol $d^2 B$ stands for the second differential form of B, namely, for the quadratic form $(H_B\, dx, dx)$.

1

Examples of Exact Bellman Functions

1.1 A Toy Problem

Let us start by considering the following simple problem. Suppose we have two positive functions f_1 and f_2 on an interval I, $I \subset \mathbb{R}$, bounded, say, by 1 and having prescribed averages: $\langle f_i \rangle_I = x_i$. We are interested in their scalar product: how large or how small it can be. That is, we would like to find the following two functions:

$$\mathbf{B}^{\max}(x_1, x_2) \stackrel{\text{def}}{=} \sup \{\langle f_1 f_2 \rangle_I : 0 \le f_i \le 1,\ \langle f_i \rangle_I = x_i\} \tag{1.1.1}$$

$$\mathbf{B}^{\min}(x_1, x_2) \stackrel{\text{def}}{=} \inf \{\langle f_1 f_2 \rangle_I : 0 \le f_i \le 1,\ \langle f_i \rangle_I = x_i\} \tag{1.1.2}$$

These functions will be called the Bellman functions of the corresponding extremal problem. In this simple case, the functions can be found by elementary consideration without using any special techniques. Nevertheless, we approach this problem as "a serious one" and provide all the steps in its derivation that we will need in the future consideration of more serious problems.

In what follows, we will consider only the first of these functions, and it will be denoted simply by $\mathbf{B}$ rather than $\mathbf{B}^{\max}$. The first question is about the domain of definition of our function. It is natural to define it on the set of all $x = (x_1, x_2) \in \mathbb{R}^2$ for which there exists at least one pair of test functions f_1 and f_2 such that $\langle f_i \rangle_I = x_i$.

DEFINITION 1.1.1 For a pair of functions $\{f_1, f_2\}$ from $L^1(I)$, we call the point $\mathfrak{b}_{f_1,f_2} \in \mathbb{R}^2$,

$$\mathfrak{b} = \mathfrak{b}_I(f_1, f_2) \stackrel{\text{def}}{=} (\langle f_1 \rangle_I, \langle f_2 \rangle_I),$$

the *Bellman point* of this pair. Very often, the pair of functions is fixed and we are interested in the dependence of the Bellman point on the interval. Then we omit arguments and use only the interval as the index:

$$\mathfrak{b}_J = (\langle f_1\rangle_{_J}, \langle f_2\rangle_{_J}) \quad \text{for any interval } J,\ J \subset I\,.$$

Clearly, the Bellman points of all admissible pairs fill the square

$$\Omega = \{x = (x_1, x_2) \colon 0 \le x_i \le 1\}\,.$$

Of course, function $\mathbf{B}$ is formally defined outside the square Ω as well, but it is not interesting to consider this function there because the supremum of the empty set is $-\infty$. Let us state this assertion as a formal proposition. It is trivial in this case, but it might not be so trivial for a more serious problem.

Proposition 1.1.2 (Domain of Definition) *The function* $\mathbf{B}$ *is defined on the domain* Ω*.*

Proof On the one hand, for any pair of test functions f_1, f_2, we have $0 \le \langle f_i\rangle_{_I} \le 1$, i.e., $\mathfrak{b}_I(f_1, f_2) \in \Omega$. On the other hand, for any $x \in \Omega$, the pair of constant functions $f_i \equiv x_i$ is an admissible pair and $\mathfrak{b}_I(x_1, x_2) = x$. □

Proposition 1.1.3 (Independence on the Interval) *The function* $\mathbf{B}$ *does not depend on the interval* I*, where the test functions are defined.*

Proof Indeed, if we have two intervals I_1 and I_2, then the linear change of variables maps the set of test functions from one interval to another preserving all averages. Therefore, for both intervals, the supremum in the definition of the Bellman function is taken over by the same set. □

We know the values of our function on the boundary $\partial\Omega$.

Proposition 1.1.4 (Boundary Conditions)

$$\begin{aligned} \mathbf{B}(0, x_2) &= 0, \qquad & \mathbf{B}(1, x_2) &= x_2, \\ \mathbf{B}(x_1, 0) &= 0, \qquad & \mathbf{B}(x_1, 1) &= x_1. \end{aligned} \tag{1.1.3}$$

Proof We easily know the boundary values because for these points, the set, over which supremum in the definition of the Bellman function is taken, consists of only one element. Indeed, if $\langle f_i\rangle_{_I} = 0$, then $f_i = 0$ almost everywhere (because $f_i \ge 0$), and therefore, $\langle f_1 f_2\rangle_{_I} = 0$. If $\langle f_i\rangle_{_I} = 1$, then $f_i = 1$ almost everywhere (because $f_i \le 1$), and hence, $\langle f_i f_j\rangle_{_I} = x_j$. □

Our function possesses an additional symmetry property:

PROPOSITION 1.1.5 (Symmetry)

$$\mathbf{B}(x_1, x_2) = \mathbf{B}(x_2, x_1)\,. \tag{1.1.4}$$

PROOF We can interchange the roles of f_1 and f_2 without changing the value of $\langle f_1 f_2 \rangle_{_I}$. Then we interchange x_1 and x_2 keeping the value of the Bellman function stable. □

PROPOSITION 1.1.6 (Main Inequality) *For every pair of points $x^\pm$ from Ω and every pair of positive numbers $\alpha^\pm$ such that $\alpha^- + \alpha^+ = 1$, the following inequality holds:*

$$\mathbf{B}(\alpha^- x^- + \alpha^+ x^+) \geq \alpha^- \mathbf{B}(x^-) + \alpha^+ \mathbf{B}(x^+). \tag{1.1.5}$$

PROOF Let us split the interval I into two parts: $I = I^- \cup I^+$ such that $|I^\pm| = \alpha^\pm |I|$. The integral in the definition of $\mathbf{B}$ can be presented as a sum of two integrals, the first over I^- and the second over I^+:

$$\int_I f_1(s) f_2(s)\, ds = \int_{I^-} f_1(s) f_2(s)\, ds + \int_{I^+} f_1(s) f_2(s)\, ds.$$

After dividing over $|I|$ we get

$$\langle f_1 f_2 \rangle_{_I} = \alpha^- \langle f_1 f_2 \rangle_{_{I^-}} + \alpha^+ \langle f_1 f_2 \rangle_{_{I^+}}.$$

Now, using the independence of the Bellman function on the interval (Proposition 1.1.3), we choose functions $f_i^\pm$ on the intervals $I^\pm$ such that they almost give us the supremum in the definition of $\mathbf{B}(x^\pm)$, i.e.,

$$\langle f_1^\pm f_2^\pm \rangle_{_{I^\pm}} \geq \mathbf{B}(x^\pm) - \eta,$$

for a fixed small $\eta > 0$. Then for the functions $f_i(s),\;\; i = 1, 2$, on I, defined as f_i^+ on I^+ and f_i^- on I^-, we obtain the inequality

$$\langle f_1 f_2 \rangle_{_I} \geq \alpha^- \mathbf{B}(x^-) + \alpha^+ \mathbf{B}(x^+) - \eta. \tag{1.1.6}$$

Observe that the pair of the compounded functions f_i is an admissible pair of test function corresponding to the point $x = \alpha^- x^- + \alpha^+ x^+$. Indeed, $x^\pm = \mathfrak{b}_{I^\pm}(f_1^\pm, f_2^\pm) = \mathfrak{b}_{I^\pm}(f_1, f_2)$, and therefore,

$$\mathfrak{b}_I(f_1, f_2) = \alpha^- \mathfrak{b}_{I^-}(f_1, f_2) + \alpha^+ \mathfrak{b}_{I^+}(f_1, f_2) = \alpha^- x^- + \alpha^+ x^+ = x.$$

The inequality $0 \leq f_i \leq 1$ is clearly fulfilled as well. So, we can take supremum in (1.1.6) over all admissible pairs of functions. This yields

$$\mathbf{B}(x) \geq \alpha^- \mathbf{B}(x^-) + \alpha^+ \mathbf{B}(x^+) - \eta,$$

which proves the main inequality because η is arbitrarily small. □

PROPOSITION 1.1.7 (Obstacle Condition)

$$\mathbf{B}(x) \geq x_1 \cdot x_2. \tag{1.1.7}$$

PROOF Since the constant functions $f_i = x_i$ belong to the set of admissible test functions corresponding to the point x, we come to the desired inequality $\sup\{\langle f_1 f_2 \rangle : \langle f_i \rangle_I = x_i\} \geq \langle x_1 x_2 \rangle_I = x_1 x_2$. □

Before stating the next proposition, we introduce some notation. Let $\mathcal{I}$ be a family of subintervals of an interval I with the following properties:

- $I \in \mathcal{I}$;
- if $J \in \mathcal{I}$, then there is a couple of almost disjoint intervals $J^{\pm}$ (i.e., with the disjoint interiors), such that $J = J^- \cup J^+$;
- $\mathcal{I} = \cup_{n\geq 0} \mathcal{I}_n$, where $\mathcal{I}_0 = \{I\}$, $\mathcal{I}_{n+1} = \{J^-, J^+ : J \in \mathcal{I}_n\}$;
- $\lim_{n\to\infty} \max\{|J| : J \in \mathcal{I}_n\} = 0$.

If the family $\mathcal{I}$ satisfies the following additional condition

- $|J^-| = |J^+|$,

it is called *dyadic*. For the dyadic family of subintervals, we use notation $\mathcal{D}(I)$ instead of $\mathcal{I}$.

PROPOSITION 1.1.8 (Bellman Induction) *If B is a continuous function on the domain Ω satisfying the main inequality* (*that is just concavity condition*) *and obstacle condition* (1.1.7), *then* $\mathbf{B}(x) \leq B(x)$.

PROOF Fix an interval I and its splitting $\mathcal{I}$. Take an arbitrary point $x \in \Omega$ and two test function f_1 and f_2 on I, $0 \leq f_i \leq 1$, such that $x = \mathfrak{b}_I(f_1, f_2)$. We can rewrite the main inequality in the form

$$|J|B(\mathfrak{b}_J)| \geq |J^+|B(\mathfrak{b}_{J^+}) + |J^-|B(\mathfrak{b}_{J^-}).$$

Let us take the sum of the earlier inequalities when J runs over $\mathcal{I}_k$, the set of subintervals of kth generation. Then $J^{\pm}$ are all intervals of the set $\mathcal{I}_{k+1}$, and we get

$$\sum_{J\in\mathcal{I}_k} |J|B(\mathfrak{b}_J) \geq \sum_{J\in\mathcal{I}_{k+1}} |J|B(\mathfrak{b}_J).$$

Therefore,

$$|I|B(x)| = |I|B(\mathfrak{b}_I) = \sum_{J\in\mathcal{I}_0} |J|B(\mathfrak{b}_J) \geq \sum_{J\in\mathcal{I}_n} |J|B(\mathfrak{b}_J) = \int_I B(x^{(n)}(s))\, ds,$$

where $x^{(n)}$ is a step function defined in the following way: $x^{(n)}(s) = \mathfrak{b}_J$, when $s \in J$, $J \in \mathcal{I}_n$.

We know that $x^{(n)}(s) \to (f_1(s), f_2(s))$ almost everywhere by the Lebesgue differentiation theorem. Since B is continuous, we have $B(x^{(n)}(s)) \to B(f_1(s), f_2(s))$. Now, using the obstacle condition (1.1.7) and the Lebesgue dominated convergence theorem, we can pass to the limit in the obtained inequality as $n \to \infty$.

$$|I|B(x) \geq \int_I B(f_1(s), f_2(s))\, ds \geq \int_I f_1(s)f_2(s)\, ds = |I|\langle f_1 f_2\rangle_I. \quad (1.1.8)$$

Taking supremum in this inequality over all admissible pairs f_1, f_2 with $\mathfrak{b}_I(f_1, f_2) = x$, we come to the desired estimate. □

According to this proposition, every concave function satisfying the obstacle condition gives us an upper estimate of the functional under consideration. If we are interested in a sharp estimate, we need to look for minimal possible such functions. Due to the symmetry (see Proposition 1.1.5), it is sufficient to consider $x_1 \leq x_2$.

On a triangle, we know our function at the vertices: $\mathbf{B}(0,0) = 0$, $\mathbf{B}(0,1) = 0$, and $\mathbf{B}(1,1) = 1$. The minimal possible concave function passing through the given three points is a linear function. In our case, it is the function $B(x) = x_1$. By the symmetry on the whole square Ω, we get the following *Bellman candidate*[1] $B(x) = \min\{x_1, x_2\}$.

In fact, we have already found the Bellman function.

THEOREM 1.1.9

$$\mathbf{B}(x) = \min\{x_1, x_2\}.$$

PROOF First of all, by Proposition 1.1.8, the upper estimate $\mathbf{B}(x) \leq B(x)$ is true because B is concave and $\min\{x_1, x_2\} \geq x_1 x_2$. Since there is no concave function satisfying the required boundary condition and that is less than B, we get $\mathbf{B} = B$.

However, in a more difficult problem, it is not so clear that the Bellman candidate cannot be diminished. By this reason, we demonstrate on this example how we will typically prove the lower estimate $\mathbf{B}(x) \geq B(x)$. To this end for every point $x \in \Omega$, we present an admissible test function, realizing the supremum in the definition of the Bellman function. In some papers, such a function (in our case, it is a pair of functions) is called an *extremizer*, but in other papers it is called an *optimizer*. We shall use both these words as synonyms. In our case, the possible pair of extremizers is very

[1] Such a term is used for a function possessing the necessary properties of the Bellman function, e.g., concavity, symmetry, boundary values, etc. After a Bellman candidate is presented, we need to check that it indeed is the desired Bellman function.

simple: $f_i = \mathbf{1}_{[0,x_i]}$. We evidently have $\langle f_i \rangle_{[0,1]} = x_i$ and $\langle f_1 f_2 \rangle_{[0,1]} = \min\{x_i\}$. Since by definition $\mathbf{B}(x)$ is the supremum of $\langle f_1 f_2 \rangle_{[0,1]}$, when f_i runs over all admissible pairs corresponding to the point x, $\mathbf{B}(x)$ is not less that this particular value, which is equal to $\min\{x_i\} = B(x)$. □

At the end of this section, we would like to explain how to find the extremizers mentioned earlier. Look at the proof of Proposition 1.1.8. Let us take $B = \mathbf{B}$ in this chain of inequalities choosing at the beginning f_1, f_2 to be a pair of extremizers. Since the first and the last terms in the chain of inequalities (1.1.8) are equal, namely, they are $|I|\mathbf{B}(x)$, we must have equalities in each step. In other words, we need to choose such a splitting $x = \alpha^- x^- + \alpha^+ x^+$ to have equality rather than inequality in (1.1.5). In our case, it is easy to do because our Bellman candidate is a concatenation of two linear functions, and if we deal only with one of these linear functions, we always have equality in (1.1.5). Based on this reason, in this simple situation, we can choose extremizers in an almost arbitrary way; the only condition is that all three points x and $x^\pm$ must be in the same triangle: either in $\{x\colon x_1 \le x_2\}$ or in $\{x\colon x_1 \ge x_2\}$.

Let us construct a pair of optimizers for some point x with $x_1 \le x_2$. First we draw the straight line passing through the points x and $x^- \stackrel{\text{def}}{=} (1,1)$. It intersects the boundary of Ω at the point $(0, \frac{x_2-x_1}{1-x_1}) \stackrel{\text{def}}{=} x^+$. So, we have $x = x_1 \cdot x^- + (1-x_1)\cdot x^+$, i.e., $\alpha^- = x_1$, $\alpha^+ = 1-x_1$, and we need to split our initial interval I (take $I=[0,1]$) in the union $I^- = [0,x_1]$ and $I^+ = [x_1,1]$. The point $x^- = (1,1)$ is the Bellman point of the only pair $f_1 = f_2 = 1$, hence on $[0,x_1]$ we take both extremal functions equal identically to 1. The point $x^+ = (0, \frac{x_2-x_1}{1-x_1})$ is the Bellman point, for example, the pair of constant functions, and we can put $f_1 = 0$ and $f_2 = \frac{x_2-x_1}{1-x_1}$ on $[x_1,1]$. It is easy to check whether this pair of functions gives us an extremizer. However, the second function of this extremizer differs from that presented earlier. What to do to get that extremizer? We only have to split I^+ once more, presenting x^+ as the convex combination of $(0,1)$ and $(0,0)$:

$$x^+ = \frac{x_2-x_1}{1-x_1}(0,1) + \frac{1-x_2}{1-x_1}(0,0), \qquad I^+ = [x_1,1] = [x_1,x_2] \cup [x_2,1].$$

The function f_1 is, as before, the zero function on both subintervals, but we have to take f_2 equal to 1 on $[x_1,x_2]$ and equal to 0 on $[x_2,1]$. In this way, we come to the pair of functions presented earlier.

We would like to provide support now to the readers for whom the latter paragraph remains unclear: you meet such kind of construction (splitting the interval and representing a Bellman point as a convex combination of two (or more) other Bellman points) many times on the pages of this book. We hope that after several repetitions, the construction becomes absolutely clear.

Exercises

PROBLEM 1.1.1 Find the function $\mathbf{B}$ defined for a similar problem, where the restriction $0 \le f_i \le 1$ is replaced by $|f_i| \le 1$

PROBLEM 1.1.2 Find the function $\mathbf{B}^{\min}$ defined in (1.1.2).

PROBLEM 1.1.3 Find the function $\mathbf{B}^{\min}$ for the set of test functions described in Problem 1.1.1.

1.2 Buckley Inequality

For an interval I and a number $r > 1$, the symbol $A_\infty(I,r)$ denotes the r-"ball" in the Muckenhoupt class A_∞:

$$A_\infty(I,r) \stackrel{\text{def}}{=} \Big\{w\colon w \in L^1(I),\ w \ge 0,\ \langle w\rangle_J \le r e^{\langle \log w\rangle_J}\ \forall J \subset I\Big\}. \tag{1.2.1}$$

We denote by $\mathcal{D}(I)$ the set of all dyadic subintervals of I and by $A_\infty^d(I,r)$ the dyadic analog of (1.2.1), i.e., in the definition of $A_\infty^d(I,r)$, we consider only $J \in \mathcal{D}(I)$.

THEOREM (Buckley [19]) *There exists a constant $c = c(r)$ such that*

$$\sum_{J\in\mathcal{D}(I)} |J|\Big(\frac{\langle w\rangle_{J+} - \langle w\rangle_{J-}}{\langle w\rangle_J}\Big)^2 \le c(r)|I|$$

for any weight w from $A_\infty^d(I,r)$.

Now, we are ready to introduce the main object of our consideration, the so-called Bellman function of the problem.

$$\begin{aligned}\mathbf{B}(x) &= \mathbf{B}(x_1,x_2;r)\\ &\stackrel{\text{def}}{=} \sup_{w\in A_\infty^d(I,r)} \left\{\frac{1}{|I|}\sum_{J\in\mathcal{D}(I)}|J|\left(\frac{\langle w\rangle_{J+}-\langle w\rangle_{J-}}{\langle w\rangle_J}\right)^2 :\right.\\ &\qquad\qquad \left.\langle w\rangle_I = x_1,\ \langle \log w\rangle_I = x_2\right\}.\end{aligned} \tag{1.2.2}$$

Let us note that we did not assign the index I to $\mathbf{B}$ despite the fact that all test functions w in its definition are considered on I. This omission is not due to our desire to simplify notation, but rather an indication of the very important fact that the function $\mathbf{B}$ does not depend on I; Proposition 1.1.3 holds in this situation by the same reason.

For a given weight $w \in A_\infty^d(I,r)$, we introduce a Bellman point $\mathfrak{b}_I(w)$ in the following way: $\mathfrak{b}_I(w) = (\langle w \rangle_I, \langle \log w \rangle_I)$. Note that for all admissible weights and for any dyadic subinterval $J \subset I$, the corresponding Bellman point $\mathfrak{b}_J(w)$ is in the following domain Ω_r:

$$\Omega_r \stackrel{\text{def}}{=} \left\{ x = (x_1, x_2) \colon \; \log \frac{x_1}{r} \le x_2 \le \log x_1 \right\}.$$

Indeed, the right bound is simply Jensen's inequality and the left one is fulfilled because our weight w is from $A_\infty^d(I,r)$.

To show that Ω_r is the domain of the function $\mathbf{B}$, we need to check that for any point $x \in \Omega_r$ there exists an admissible weight with $\mathfrak{b}_I(w) = x$. However, we leave this for the reader as an exercise (see Problem 1.2.1).

Now we prove the crucial property of the function $\mathbf{B}$ that follows directly from its definition.

Lemma 1.2.1 (Main Inequality) *For every pair of points $x^\pm$ from Ω_r such that their mean $x = (x^+ + x^-)/2$ is also in Ω_r, the following inequality holds:*

$$\mathbf{B}(x) \ge \frac{\mathbf{B}(x^+) + \mathbf{B}(x^-)}{2} + \left(\frac{x_1^+ - x_1^-}{x_1} \right)^2. \qquad (1.2.3)$$

Proof Let us split the sum in the definition of $\mathbf{B}$ into three parts: the sum over $\mathcal{D}(I^+)$, the sum over $\mathcal{D}(I^-)$, and an additional term corresponding to I itself:

$$\begin{aligned} &\frac{1}{|I|} \sum_{J \in \mathcal{D}(I)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2 \\ &\quad = \frac{1}{2|I^+|} \sum_{J \in \mathcal{D}(I^+)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2 \\ &\qquad + \frac{1}{2|I^-|} \sum_{J \in \mathcal{D}(I^-)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_I} \right)^2 \\ &\qquad + \left(\frac{\langle w \rangle_{I^+} - \langle w \rangle_{I^-}}{\langle w \rangle_I} \right)^2. \end{aligned}$$

Using the fact that $\mathbf{B}$ does not depend on the interval where the test functions are defined, we can choose two weights $w^\pm$ on the intervals $I^\pm$ that almost give us the supremum in the definition of $\mathbf{B}(x^\pm)$, i.e.,

$$\frac{1}{|I^\pm|}\sum_{J\in\mathcal{D}(I^\pm)}|J|\left(\frac{\langle w^\pm\rangle_{J+}-\langle w^\pm\rangle_{J-}}{\langle w^\pm\rangle_J}\right)^2\geq \mathbf{B}(x^\pm)-\eta,$$

for an arbitrary fixed small $\eta>0$. Then for the weight w on I, defined as w^+ on I^+ and w^- on I^-, we obtain the inequality

$$\frac{1}{|I|}\sum_{J\in\mathcal{D}(I)}|J|\left(\frac{\langle w\rangle_{J+}-\langle w\rangle_{J-}}{\langle w\rangle_J}\right)^2\geq \frac{\mathbf{B}(x^+)+\mathbf{B}(x^-)}{2}-\eta+\left(\frac{x_1^+-x_1^-}{x_1}\right)^2. \tag{1.2.4}$$

Observe that the compound weight w is an admissible weight, corresponding to the point x. Indeed, $x^\pm=\mathfrak{b}_{I^\pm}(w)$ and by the construction of $w^\pm$ we have $w^\pm\in A_\infty^d(I^\pm,r)$. Therefore, the weight w satisfies the inequality $\langle w\rangle_J\leq re^{\langle\log w\rangle_J}$ for all $J\in\mathcal{D}(I^+)$, since w^+ does, and for all $J\in\mathcal{D}(I^-)$, since w^- does. Lastly, $\langle w\rangle_I\leq re^{\langle\log w\rangle_I}$, because, by assumption, $x\in\Omega_r$.

We can now take supremum in (1.2.4) over all admissible weights w, which yields

$$\mathbf{B}(x)\geq\frac{\mathbf{B}(x^+)+\mathbf{B}(x^-)}{2}-\eta+\left(\frac{x_1^+-x_1^-}{x_1}\right)^2.$$

This proves the main inequality because η is arbitrarily small. □

Lemma 1.2.2 (Boundary Condition)

$$\mathbf{B}(x_1,\log x_1)=0.$$

Proof Let us take a boundary point x of our domain Ω_r, that is a point with $x_2=\log x_1$. Since the equality in Jensen's inequality $e^{\langle w\rangle}\leq\langle e^w\rangle$ occurs only for constant functions w, the only test function corresponding to x is the constant (up to a set of measure zero) weight $w=x_1$. So, on this boundary, we have $\mathbf{B}(x)=0$. □

Lemma 1.2.3 (Homogeneity) *There is a function g on $[1,r]$ satisfying $g(1)=0$ and such that*

$$\mathbf{B}(x)=\mathbf{B}(x_1e^{-x_2},0)=g(x_1e^{-x_2}).$$

Proof For a weight w on an interval I and a positive number τ, consider a new weight $\tilde{w}=\tau w$. If $x=\mathfrak{b}_I(w)$, i.e., $x_1=\langle w\rangle_I,\ \ x_2=\langle\log w\rangle_I$, then for the point $\mathfrak{b}_I(\tilde{w})=\tilde{x}$ we have $\tilde{x}_1=\tau x_1$, $\tilde{x}_2=x_2+\log\tau$. Note that the expression in the definition of $\mathbf{B}$ is homogeneous of order 0 with respect to w, i.e., it does not depend on τ. Since the weights w and $\tilde{w}$ run over the whole set $A_\infty^d(I,r)$ simultaneously, we get $\mathbf{B}(x) = \mathbf{B}(\tilde{x})$. Choosing $\tau = e^{-x_2}$, we obtain

$$\mathbf{B}(x) = \mathbf{B}(x_1 e^{-x_2}, 0).$$

To complete the proof, it suffices to take $g(s) = \mathbf{B}(s,0)$. The boundary condition $g(1) = 0$ holds due to Lemma 1.2.2. □

We are now ready to demonstrate how the Bellman induction works in this case.

Lemma 1.2.4 (Bellman Induction) *Let B be a nonnegative function on Ω_r satisfying the main inequality in Ω_r (Lemma 1.2.1). Then*

$$\mathbf{B}(x) \le B(x).$$

Proof Fix an interval I and a point $x \in \Omega_r$. Take an arbitrary weight $w \in A_\infty^d(I,r)$ such that $\mathfrak{b}_I(w) = x$. Let us repeatedly use the main inequality in the form

$$|J|\, B(\mathfrak{b}_J) \ge |J^+|\, B(\mathfrak{b}_{J^+}) + |J^-|\, B(\mathfrak{b}_{J^-}) + |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2,$$

applying it first to I, then to the intervals of the first generation (that is $I^\pm$), and so on until $\mathcal{D}_n(I)$:

$$\begin{aligned} |I|\, B(\mathfrak{b}_I) &\ge |I^+|\, B(\mathfrak{b}_{I^+}) + |I^-|\, B(\mathfrak{b}_{I^-}) + |I| \left(\frac{\langle w \rangle_{I^+} - \langle w \rangle_{I^-}}{\langle w \rangle_I} \right)^2 \\ &\ge \sum_{J \in \mathcal{D}_n(I)} |J|\, B(\mathfrak{b}_J) + \sum_{k=0}^{n-1} \sum_{J \in \mathcal{D}_k(I)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2. \end{aligned}$$

Therefore,

$$\sum_{k=0}^{n-1} \sum_{J \in \mathcal{D}_k(I)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2 \le |I|\, B(\mathfrak{b}_I),$$

and passing to the limit as $n \to \infty$, we get

$$\sum_{J \in \mathcal{D}(I)} |J| \left(\frac{\langle w \rangle_{J^+} - \langle w \rangle_{J^-}}{\langle w \rangle_J} \right)^2 \le |I|\, B(x).$$

Taking supremum over all admissible weight w corresponding to the point x, we come to the desired estimate. □

COROLLARY 1.2.5 *Let g be a nonnegative function on $[1,r]$ such that the function $B(x) \stackrel{\text{def}}{=} g(x_1e^{-x_2})$ satisfies inequality (1.2.3) in Ω_r. Then Buckley's inequality holds with the constant $c(r) = \|g\|_{L^\infty([1,r])}$.*

A natural question arises: How to find such a function g? To answer it, we first replace our main inequality, which is an inequality in finite differences, by a differential inequality. Let us denote the difference between x^+ and x^- by 2Δ, then $x^\pm = x \pm \Delta$ and the Taylor expansion around the point x gives us

$$\begin{aligned} B(x^\pm) &= B(x) \pm \frac{\partial B}{\partial x_1}\Delta_1 \pm \frac{\partial B}{\partial x_2}\Delta_2 \\ &\qquad + \frac{1}{2}\frac{\partial^2 B}{\partial x_1^2}\Delta_1^2 + \frac{\partial^2 B}{\partial x_1 \partial x_2}\Delta_1\Delta_2 + \frac{1}{2}\frac{\partial^2 B}{\partial x_2^2}\Delta_2^2 + o(|\Delta|^2), \end{aligned}$$

and, therefore,

$$\begin{aligned} &\frac{B(x^+) + B(x^-)}{2} + \left(\frac{x_1^+ - x_1^-}{x_1}\right)^2 - B(x) \\ &\quad = \frac{1}{2}\frac{\partial^2 B}{\partial x_1^2}\Delta_1^2 + \frac{\partial^2 B}{\partial x_1 \partial x_2}\Delta_1\Delta_2 + \frac{1}{2}\frac{\partial^2 B}{\partial x_2^2}\Delta_2^2 + 4\left(\frac{\Delta_1}{x_1}\right)^2 + o(|\Delta|^2). \end{aligned}$$

Thus, under the assumption that our candidate B is sufficiently smooth, the main inequality (1.2.3) implies the following matrix differential inequality:

$$\begin{pmatrix} \dfrac{\partial^2 B}{\partial x_1^2} + \dfrac{8}{x_1^2} & \dfrac{\partial^2 B}{\partial x_1 \partial x_2} \\[2ex] \dfrac{\partial^2 B}{\partial x_1 \partial x_2} & \dfrac{\partial^2 B}{\partial x_2^2} \end{pmatrix} \le 0. \tag{1.2.5}$$

That is, this matrix has to be nonpositively defined.

By the preceding two lemmata, we can restrict our search to functions B of the form $B(x_1, x_2) = g(x_1e^{-x_2})$, where g is a function on the interval $[1,r]$. In terms of g, our condition (1.2.5) can be rewritten as follows:

$$\begin{pmatrix} e^{-2x_2}\left(g'' + \dfrac{8}{s^2}\right) & -e^{-x_2}(sg')' \\[2ex] -e^{-x_2}(sg')' & s(sg')' \end{pmatrix} \le 0,$$

where $g = g(s)$ and $s = x_1e^{-x_2}$. From this matrix inequality, we conclude that

$$g'' + \frac{8}{s^2} \le 0, \tag{1.2.6}$$

$$(sg')' \le 0, \tag{1.2.7}$$

and that the determinant of the matrix must be nonnegative. However, we replace the last requirement by a stronger one – we require the determinant to be identically zero. This requirement comes from our desire to find the best possible estimate: If we take an extremal weight w, i.e., a weight on which the supremum in the definition of the Bellman function is attained, then we must have equalities on each step of the Bellman induction; therefore, on each step the main inequality (1.2.3) becomes an equality. Thus, for each dyadic subinterval J of I, there exists a direction through the point $\mathfrak{b}_J$ in Ω_r along which the quadratic form given by (1.2.5) is identically zero. Hence, the matrix (1.2.5) has a nontrivial kernel and so must have a zero determinant.[2]

Calculating the determinant, we get the equation

$$\Big(g' - \frac{8}{s}\Big)(sg')' = 0.$$

The general solution of this equation is $g(s) = c\log s + c_1$. Due to the boundary condition $g(1) = 0$, we have to take $c_1 = 0$.

Now we need to choose another constant, c. To this end, we return to the necessary conditions (1.2.6–1.2.7). The second inequality is fulfilled for all c because the expression is identically zero, while the first one gives $c \geq 8$. Since we would like to have g as small as possible (as it gives the upper bound in Buckley's inequality), it is natural to take $c = 8$. Finally, we get

$$g(s) = 8\log s \qquad \text{and} \qquad B(x_1, x_2) = 8(\log x_1 - x_2).$$

Lemma 1.2.6 *The function*

$$B(x_1, x_2) = 8(\log x_1 - x_2)$$

satisfies the main inequality (1.2.3).

Proof Put, as before, $\Delta = \frac{1}{2}(x^+ - x^-)$, so $x^\pm = x \pm \Delta$. Then

$$\begin{aligned} B(x) &- \frac{B(x^+) + B(x^-)}{2} - \left(\frac{x_1^+ - x_1^-}{x_1}\right)^2 \\ &= 8\log x_1 - 8x_2 - 4\log(x_1^+ x_1^-) + 4(x_2^+ + x_2^-) - \left(\frac{x_1^+ - x_1^-}{x_1}\right)^2 \end{aligned}$$

[2] This is not a proof, the arguments are not absolutely correct; for example, the existence of an extremal weight w is not guaranteed. The supremum in the definition of the Bellman function can be not attainable and only an extremal sequence of weights can realize it. (By the way, for the Buckley inequality it is just the case.) Nevertheless, in the process of searching for a Bellman candidate, we may assume whatever we want (e.g., its smoothness to replace a finite difference condition by a differential one), but a rigorous proof starts after a candidate is found and we check that it is the true Bellman function.

$$= 4\log\frac{x_1^2}{(x_1+\Delta_1)(x_1-\Delta_1)} - 4\left(\frac{\Delta_1}{x_1}\right)^2$$

$$= -4\left[\log\left(1-\left(\frac{\Delta_1}{x_1}\right)^2\right) + \left(\frac{\Delta_1}{x_1}\right)^2\right] \geq 0.$$

□

Now we can apply Lemma 1.2.4 to $g(s) = 8\log s$, which yields the following:

THEOREM *The estimate*

$$\sum_{J\in\mathcal{D}(I)} |J|\Big(\frac{\langle w\rangle_{J+} - \langle w\rangle_{J-}}{\langle w\rangle_J}\Big)^2 \leq 8\log r\,|I|$$

holds for any weight $w \in A_\infty^d(I,r)$.

We would like to emphasize that we still have not found the Bellman function $\mathbf{B}$. The theorem just proved guarantees only the estimate

$$\mathbf{B}(x) \leq 8(\log x_1 - x_2).$$

To prove that this Bellman candidate is the true Bellman function (what proves sharpness of the earlier estimate), we need to find extremizers for every point of the domain Ω_r. However, this is a much more difficult task than it was in our previous example. For this reason, we now stop our investigation of the Bellman function for the Buckley inequality. The more experienced reader interested in completing investigation of this Bellman function can refer to Section 1.10, where we not only present the extremizers for the discussed Bellman function but also find the minimal Bellman function with completely different extremizers.

Exercises

PROBLEM 1.2.1 Check that the function $\mathbf{B}$ defined on the whole domain Ω_r, i.e., for every point x, $x \in \Omega_r$, there exists a function $w \in A_\infty^d(I,r)$ such that $x = \mathfrak{b}_I(w)$.

PROBLEM 1.2.2 Try to repeat the earlier procedure for finding the function $\mathbf{B}^{\min}$ defined by (1.2.2), where sup is replaced by inf.

PROBLEM 1.2.3 Try to find the following Bellman function

$$\mathbf{B}(x;m,M) \stackrel{\text{def}}{=} \sup_{u,v}\Big\{\frac{1}{|I|}\sum_{J\in\mathcal{D}(I)} |(u,h_J)|\,|(v,h_J)|\Big\},$$

where the supremum is taken over the set of all pairs of weights $u,\ v$ such that $\langle u\rangle_I = x_1,\ \langle v\rangle_I = x_2$, and $m^2 \le \langle u\rangle_J\langle v\rangle_J \le M^2,\ \forall J\in\mathcal{D}(I)$.

As a result you have to prove the following theorem:

THEOREM *If two weights $u, v \in L^1(I)$ satisfy the condition*

$$\sup_{J\in\mathcal{D}(I)} \langle u\rangle_J\langle v\rangle_J \le M^2,$$

then

$$\frac{1}{|I|}\sum_{J\in\mathcal{D}(I)} |J|\,|\langle u\rangle_{J+} - \langle u\rangle_{J-}|\,|\langle v\rangle_{J+} - \langle v\rangle_{J-}| \le 16M\sqrt{\langle u\rangle_I\langle v\rangle_I}\,.$$

1.3 John–Nirenberg Inequality

A function $\varphi\in L^1(I)$ is said to belong to the space $\mathrm{BMO}(I)$ if

$$\sup_J \langle|\varphi(s) - \langle\varphi\rangle_J|\rangle_J < \infty$$

for all subintervals $J \subset I$. If this condition holds only for the dyadic subintervals $J\in\mathcal{D}(I)$, we will write $\varphi\in\mathrm{BMO}^d(I)$. In fact, the following is true for any $p,\ p\in(0,\infty)$:

$$\varphi\in\mathrm{BMO}(I) \iff \Big(\sup_{J\subset I}\frac{1}{|J|}\int_J |\varphi(s)-\langle\varphi\rangle_J|^p\,ds\Big)^{\frac{1}{p}} < \infty.$$

If we factor over the constants, we get a normed space, where the expression on the right-hand side can be taken as one of the equivalent norm for any $p\in[1,\infty)$. In what follows, we will use the L^2-based norm:

$$\|\varphi\|^2_{\mathrm{BMO}(I)} = \sup_{J\subset I}\frac{1}{|J|}\int_J |\varphi(s)-\langle\varphi\rangle_J|^2\,ds = \sup_{J\subset I}\left(\langle\varphi^2\rangle_J - \langle\varphi\rangle_J^2\right).$$

The BMO ball of radius ε centered at 0 will be denoted by BMO_ε. Using the Haar decomposition

$$\varphi(s) = \langle\varphi\rangle_I + \sum_{J\in\mathcal{D}(I)} (\varphi, h_J)h_J(s),$$

we can write down the expression for the norm in the following way:

$$\begin{aligned}\|\varphi\|^2_{\mathrm{BMO}(I)} &= \sup_{J\subset I}\frac{1}{|J|}\sum_{L\in\mathcal{D}(J)} |(\varphi, h_L)|^2\\ &= \frac{1}{4}\sup_{J\subset I}\frac{1}{|J|}\sum_{L\in\mathcal{D}(J)} |L|\left(\langle\varphi\rangle_{L+} - \langle\varphi\rangle_{L-}\right)^2.\end{aligned}$$

THEOREM (John–Nirenberg [88]) *There exist absolute constants c_1 and c_2 such that*

$$\bigl|\{s\in I\colon |\varphi(s)-\langle\varphi\rangle_{I}|\ge\lambda\}\bigr|\le c_1 e^{-c_2\frac{\lambda}{\|\varphi\|}}|I|$$

for all $\varphi\in\mathrm{BMO}_\varepsilon(I)$.

An equivalent, integral form of the same assertion is the following:

THEOREM *There exists an absolute constant ε_0 such that for any $\varphi\in\mathrm{BMO}_\varepsilon(I)$ with $\varepsilon<\varepsilon_0$, the inequality*

$$\langle e^{\varphi}\rangle_{I}\le c\,e^{\langle\varphi\rangle_I}$$

holds with a constant $c=c(\varepsilon)$ not depending on φ.

We shall prove the theorem in this integral form and find the sharp constant $c(\varepsilon)$. Our Bellman function

$$\mathbf{B}(x;\varepsilon)\stackrel{\text{def}}{=}\sup_{\varphi\in\mathrm{BMO}_\varepsilon(I)}\left\{\langle e^{\varphi}\rangle_{I}\colon \mathfrak{b}_I(\varphi)=x\right\},$$

where $\mathfrak{b}_I(\varphi)\stackrel{\text{def}}{=}(\langle\varphi\rangle_{I},\langle\varphi^2\rangle_{I})$ is the Bellman point corresponding to the test function φ and the interval I. It is clear that the set of all Bellman points is the domain

$$\Omega_\varepsilon\stackrel{\text{def}}{=}\left\{x=(x_1,x_2)\colon x_1^2\le x_2\le x_1^2+\varepsilon^2\right\},$$

i.e., Ω_ε is the domain where $\mathbf{B}$ is defined. Let us note from the beginning that we will consider $\varepsilon<1$ only because $\varphi(s)=-\log s\in\mathrm{BMO}_1([0,1])$ and $\langle e^{\varphi}\rangle_{[0,1]}=\infty$.

First, we will consider the dyadic problem and deduce the main inequality for the dyadic Bellman function.

LEMMA 1.3.1 (Main Inequality) *For every pair of points $x^\pm$ from Ω_ε such that their mean $x=(x^++x^-)/2$ is also in Ω_ε, the following inequality holds*

$$\mathbf{B}(x)\ge\frac{\mathbf{B}(x^+)+\mathbf{B}(x^-)}{2}.\tag{1.3.1}$$

PROOF The proof repeats almost verbatim the proof of the main inequality for the Buckley's Bellman function. We split the integral in the definition of $\mathbf{B}$ into two parts, the integral over I^+ and the one over I^- :

$$\int_I e^{\varphi(s)}\,ds=\int_{I^+}e^{\varphi(s)}\,ds+\int_{I^-}e^{\varphi(s)}\,ds.$$

Now we choose such functions $\varphi^{\pm}$ on the intervals $I^{\pm}$ that they almost give us the supremum in the definition of $\mathbf{B}(x^{\pm})$, i.e,

$$\frac{1}{|I^{\pm}|}\int_{I^{\pm}} e^{\varphi^{\pm}(s)}\,ds \geq \mathbf{B}(x^{\pm}) - \eta,$$

for a fixed small $\eta > 0$. Then for the function φ on I, defined as φ^{+} on I^{+} and φ^{-} on I^{-}, we obtain the inequality

$$\frac{1}{|I|}\int_{I} e^{\varphi(s)}\,ds \geq \frac{\mathbf{B}(x^{+}) + \mathbf{B}(x^{-})}{2} - \eta. \tag{1.3.2}$$

Observe that the compound function φ is an admissible test function corresponding to the point x. Indeed, $x^{\pm} = \mathfrak{b}_{I^{\pm}}(\varphi)$ and by construction $\varphi^{\pm} \in \mathrm{BMO}_{\varepsilon}^{d}(I^{\pm})$; therefore, the function φ satisfies the inequality $\langle\varphi^2\rangle_{_J} - \langle\varphi\rangle_{_J}^2 \leq \varepsilon^2$ for all $J \in \mathcal{D}(I^{+})$, since φ^{+} does, and for all $J \in \mathcal{D}(I^{-})$, since φ^{-} does. Lastly, $\langle\varphi^2\rangle_{_I} - \langle\varphi\rangle_{_I}^2 \leq \varepsilon^2$, because, by assumption, $x \in \Omega_{\varepsilon}$.

We can now take supremum in (1.3.2) over all admissible functions φ, which yields

$$\mathbf{B}(x) \geq \frac{\mathbf{B}(x^{+}) + \mathbf{B}(x^{-})}{2} - \eta.$$

This proves the main inequality because η is arbitrarily small. □

As in the case of the Buckley inequality, the next step is to derive a boundary condition for $\mathbf{B}$.

LEMMA 1.3.2 (Boundary Condition)

$$\mathbf{B}(x_1, x_1^2) = e^{x_1}. \tag{1.3.3}$$

PROOF The function $\varphi(s) = x_1$ is the only test function corresponding to the point $x = (x_1, x_1^2)$, because the equality in the Hölder inequality $x_2 \geq x_1^2$ occurs only for constant functions. Hence, $e^{\varphi} = e^{x_1}$. □

Now we are ready to describe super-solutions as functions verifying the main inequality and the boundary conditions.

LEMMA 1.3.3 (Bellman Induction) *If B is a continuous function on the domain Ω_{ε}, satisfying the main inequality* (1.3.1) *for any pair $x^{\pm}$ of points from Ω_{ε} such that $x \overset{\text{def}}{=} \frac{x^{+}+x^{-}}{2} \in \Omega_{\varepsilon}$, as well as the boundary condition* (1.3.3), *then $\mathbf{B}(x) \leq B(x)$.*

PROOF Fix a bounded function $\varphi \in \mathrm{BMO}_{\varepsilon}(I)$ and put $x = \mathfrak{b}_I(\varphi)$. As in the case of Buckley inequality, we rewrite the main inequality in the form

$$|J|\, B(\mathfrak{b}_J) \geq |J^{+}|\, B(\mathfrak{b}_{J^{+}}) + |J^{-}|\, B(\mathfrak{b}_{J^{-}}),$$

applying it first to I, then to the intervals of the first generation (that is $I^{\pm}$), and so on until $\mathcal{D}_n(I)$:

$$\begin{aligned}|I|B(\mathfrak{b}_I)| &\geq |I^+|B(\mathfrak{b}_{I^+}) + |I^-|B(\mathfrak{b}_{I^-})\\ &\geq \sum_{J\in\mathcal{D}_n(I)} |J|B(\mathfrak{b}_J) = \int_I B(x^{(n)}(s))\,ds,\end{aligned}$$

where $x^{(n)}(s) = \mathfrak{b}_J$, when $s \in J$, $\ J \in \mathcal{D}_n(I)$. (Recall that $\mathcal{D}_n(I)$ stands for the set of subintervals of n-th generation.) By the Lebesgue differentiation theorem, we have $x^{(n)}(s) \to (\varphi(s), \varphi^2(s))$ almost everywhere. Now, we can pass to the limit in this inequality as $n \to \infty$. Since φ is assumed to be bounded, $x^{(n)}(s)$ runs in a bounded (and, therefore, compact) subdomain of Ω_ε. Since B is continuous, it is bounded on any compact set and so, by the Lebesgue dominated convergence theorem, we can pass to the limit in the integral using the boundary condition (1.3.3):

$$|I|B(\mathfrak{b}_I(\varphi)) \geq \int_I B(\varphi(s), \varphi^2(s))\,ds = \int_I e^{\varphi(s)}ds = |I|\langle e^{\varphi}\rangle_{I}. \tag{1.3.4}$$

To complete the proof of the lemma, we need to pass from bounded to arbitrary BMO test functions. To this end, we will use the following result:

Lemma 1.3.4 (Cut-Off Lemma) *Fix $\varphi \in \mathrm{BMO}(I)$ and two real numbers c, d such that $c < d$. Let $\varphi_{c,d}$ be the cut-off of φ at heights c and d :*

$$\varphi_{c,d}(s) = \begin{cases} c, & if\ \varphi(s) \leq c;\\ \varphi(s), & if\ c < \varphi(s) < d;\\ d, & if\ \varphi(s) \geq d.\end{cases} \tag{1.3.5}$$

Then

$$\langle \varphi_{c,d}^2\rangle_{J} - \langle\varphi_{c,d}\rangle_{J}^2 \leq \langle\varphi^2\rangle_{J} - \langle\varphi\rangle_{J}^2, \quad \forall J,\ J\subset I,$$

and, consequently,

$$\|\varphi_{c,d}\|_{\mathrm{BMO}} \leq \|\varphi\|_{\mathrm{BMO}}.$$

Proof If we integrate the evident inequality

$$|\varphi_{c,d}(s) - \varphi_{c,d}(t)|^2 \leq |\varphi(s) - \varphi(t)|^2$$

over J with respect to s and once more with respect to t, then dividing the result over $2|J|^2$, we get the desired estimate. □

Now, let $\varphi \in \mathrm{BMO}_\varepsilon(I)$ be a function bounded from above. Then, by the earlier lemma, $\varphi_n \stackrel{\mathrm{def}}{=} \varphi_{-n,\infty} \in \mathrm{BMO}_\varepsilon(I)$. For the bounded function φ_n, inequality (1.3.4) is true, i.e.,

$$B(\langle\varphi_n\rangle_{{}_I}, \langle\varphi_n^2\rangle_{{}_I}) \geq \langle e^{\varphi_n}\rangle_{{}_I}.$$

Since e^{φ_0} is a summable majorant for e^{φ_n} and B is continuous, we can pass to the limit and obtain the estimate (1.3.4) for any function φ bounded from above. Finally, we repeat this approximation procedure for an arbitrary φ. Now, we take $\varphi_n = \varphi_{-\infty,n}$ and we can pass to the limit in the right-hand side of the inequality by the monotone convergence theorem.

So, we have proved the inequality

$$B(\mathfrak{b}_I(\varphi)) \geq \langle e^{\varphi}\rangle_{{}_I}$$

for arbitrary $\varphi \in \mathrm{BMO}_\varepsilon(I)$. Taking supremum over all admissible test functions corresponding to the point x, i.e., over all φ such that $\mathfrak{b}_I(\varphi) = x$, we get $B(x) \geq \mathbf{B}(x)$. $\square$

As before, to come up with a candidate for the Bellman function, we pass from the finite difference inequality (1.3.1) to the infinitesimal one:

$$\frac{d^2B}{dx^2} \stackrel{\mathrm{def}}{=} \begin{pmatrix} \dfrac{\partial^2 B}{\partial x_1^2} & \dfrac{\partial^2 B}{\partial x_1 \partial x_2} \\[2ex] \dfrac{\partial^2 B}{\partial x_1 \partial x_2} & \dfrac{\partial^2 B}{\partial x_2^2} \end{pmatrix} \leq 0, \tag{1.3.6}$$

and we will require this Hessian matrix to be degenerate, i.e., $\det(\frac{d^2B}{dx^2}) = 0$. Again, to solve this PDE, we use a homogeneity property to reduce the problem to an ODE.

Lemma 1.3.5 (Homogeneity) *There exists a function G on the interval $[0, \varepsilon^2]$ such that*

$$\mathbf{B}(x;\varepsilon) = e^{x_1} G(x_2 - x_1^2), \qquad G(0) = 1.$$

Proof Let φ be an arbitrary test function defined on an interval I and $x = (\langle\varphi\rangle_{{}_I}, \langle\varphi^2\rangle_{{}_I})$ its Bellman point. Then the function $\tilde{\varphi} \stackrel{\mathrm{def}}{=} \varphi + \tau$ is also a test function with the same norm, and its Bellman point is

$$\tilde{x} = (x_1 + \tau, x_2 + 2\tau x_1 + \tau^2).$$

If φ runs over the set of all test functions corresponding to x, then $\tilde{\varphi}$ runs over the set of all test functions corresponding to $\tilde{x}$ and vice versa. Therefore,

$$\mathbf{B}(\tilde{x}) = \sup_{\tilde{\varphi}} \langle e^{\tilde{\varphi}} \rangle_{_I} = e^{\tau} \sup_{\varphi} \langle e^{\varphi} \rangle_{_I} = e^{\tau} \mathbf{B}(x).$$

Choosing $\tau = -x_1$, we get

$$\mathbf{B}(x) = e^{-\tau} \mathbf{B}(x_1 + \tau, x_2 + 2\tau x_1 + \tau^2) = e^{x_1} \mathbf{B}(0, x_2 - x_1^2).$$

Setting $G(s) = \mathbf{B}(0, s)$ completes the proof. □

Since $G > 0$, we can introduce $g(s) = \log G(s)$ and look for a function B of the form

$$B(x_1, x_2) = e^{x_1 + g(x_2 - x_1^2)}.$$

By direct calculation, we get

$$\begin{aligned}
\frac{\partial^2 B}{\partial x_1^2} &= \left(1 - 4x_1 g' + 4x_1^2 (g')^2 - 2g' + 4x_1^2 g''\right) B, \\
\frac{\partial^2 B}{\partial x_1 \partial x_2} &= \left(g' - 2x_1 (g')^2 - 2x_1 g''\right) B, \\
\frac{\partial^2 B}{\partial x_2^2} &= \left((g')^2 + g''\right) B.
\end{aligned} \tag{1.3.7}$$

The partial differential equation $\det(\frac{d^2 B}{dx^2}) = 0$ then turns into the following ordinary differential equation:

$$\begin{aligned}
&\left(1 - 4x_1 g' + 4x_1^2 (g')^2 - 2g' + 4x_1^2 g''\right)\left((g')^2 + g''\right) \\
&\qquad\qquad = \left(g' - 2x_1 (g')^2 - 2x_1 g''\right)^2,
\end{aligned}$$

which reduces to

$$g'' - 2g'g'' - 2(g')^3 = 0. \tag{1.3.8}$$

Dividing by $2(g')^3$ (since we are not interested in constant solutions), we get

$$\left(\frac{1}{g'} - \frac{1}{4(g')^2}\right)' = 1,$$

which yields

$$\frac{1}{g'} - \frac{1}{4(g')^2} = s + \text{const}$$

or, equivalently,

$$-\left(1 - \frac{1}{2g'}\right)^2 = s + \text{const}, \qquad \forall s \in [0, \varepsilon^2].$$

Since the left-hand side is nonpositive, the constant cannot be greater than $-\varepsilon^2$. Let us denote it by $-\delta^2$, where $\delta \geq \varepsilon$.

Thus, we have two possible solutions:

$$1 - \frac{1}{2g'_\pm} = \pm\sqrt{\delta^2 - s}. \tag{1.3.9}$$

Using the boundary condition $g(0) = 0$, we obtain

$$g_\pm(s) = \frac{1}{2}\int_0^s \frac{dt}{1 \mp \sqrt{\delta^2 - t}} = \log\frac{1 \mp \sqrt{\delta^2 - s}}{1 \mp \delta} \pm \sqrt{\delta^2 - s} \mp \delta.$$

These functions are well defined for $\delta \in [\varepsilon, 1)$ and give us two solutions for B:

$$B_\pm(x;\delta) = \frac{1 \mp \sqrt{\delta^2 - x_2 + x_1^2}}{1 \mp \delta}\exp\left\{x_1 \pm \sqrt{\delta^2 - x_2 + x_1^2} \mp \delta\right\}. \tag{1.3.10}$$

For the function $B = B_+(x;\delta)$, the Hessian $\frac{d^2B}{dx^2}$ is nonpositive and it is nonnegative for $B = B_-(x;\delta)$. This is possible to check either by the direct calculation (see Problem 1.3.1) or using formula (1.3.9) for g' and expression (1.3.7) for $B_{x_2x_2}$ together with relation (1.3.8) between g'' and g'. Therefore, we have to choose our Bellman candidate among the family of functions $B = B_+(x;\delta),\ \ \delta \in [\varepsilon, 1)$. Since

$$\frac{\partial B}{\partial \delta} = \frac{\delta^2}{(1-\delta)^2}\exp\left\{x_1 \pm \sqrt{\delta^2 - x_2 + x_1^2} \mp \delta\right\},$$

i.e., B increases in δ, and we are interested in the minimal possible majorant, it is natural to choose $\delta = \varepsilon$. However, this choice does not give us the dyadic Bellman function because for this function, the main inequality (1.3.1) is not fulfilled (see Problem 1.3.3). This is in contrast to the Buckley's Bellman function from Section 1.2.

To choose the proper δ for a given ε is not a simple task and for this reason, we refer the reader to [**171**]. Now we concentrate our attention on the function $B(x;\varepsilon)$. As already mentioned, this function is not concave in the strip Ω_ε (i.e., inequality (1.3.1) is not fulfilled everywhere in Ω_ε), but it is *locally concave*. This means that it is concave in every convex subset of Ω_ε because its Hessian is negative.

The latter statement is simple, but for the reader who is not familiar with the fact that a smooth function is locally concave in a domain if and only if its Hessian is nonpositive in this domain, we present a proof. Let us parametrize the interval $[x^-, x^+]$ as follows: $x(s) = (1-s)x^- + sx^+$,

$0 \le s \le 1$, and put $b(s) \stackrel{\text{def}}{=} B(x(s))$. Then for arbitrary $\alpha^{\pm}$, $\alpha^+ + \alpha^- = 1$, we have

$$B(\alpha^- x^- + \alpha^+ x^+) - \alpha^- B(x^-) - \alpha^+ B(x^+) \\ = b(\alpha^+) - \alpha^- b(0) - \alpha^+ b(1) = -\int_0^1 k(\alpha, s) b''(s)\, ds,$$

where

$$k(\alpha, s) = \begin{cases} \alpha^- s & \text{for} \quad 0 \le s \le \alpha^+, \\ \alpha^+(1-s) & \text{for} \quad \alpha^+ \le s \le 1. \end{cases}$$

And since

$$b''(s) = \sum_{i,j=1}^{2} B_{x_i x_j}(x(s))(x_i^+ - x_i^-)(x_j^+ - x_j^-) \le 0,$$

we have the required concavity condition

$$B(\alpha^- x^- + \alpha^+ x^+) \ge \alpha^- B(x^-) + \alpha^+ B(x^+).$$

Our next goal is to show that $B_+(x, \varepsilon)$ is the Bellman function for classical (non-dyadic) BMO. We need to modify the Bellman induction for this situation because now we have significant freedom: We need to choose how to split the interval on each step of induction. For a given test function φ, we will split the interval I not in two equal halves as for dyadic case, but try to split it in two parts $I = I^+ \cup I^-$ in such a way that the segment $[\mathfrak{b}_{I^-}, \mathfrak{b}_{I^+}]$ is entirely if not in Ω_ε then in a slightly larger domain Ω_δ. If this were possible for some δ, we could run the Bellman induction for $B(x;\delta)$ and get $\mathbf{B}(x;\varepsilon) \le B(x;\delta)$. If such a procedure were possible for every $\delta > \varepsilon$, we could pass to the limit $\delta \to \varepsilon$ to get the final upper estimate $\mathbf{B}(x;\varepsilon) \le B(x;\varepsilon)$.

To realize this plan, we prove the following purely geometric result that is crucial to applying the Bellman function method to the usual, non-dyadic BMO.

Lemma 1.3.6 (Splitting Lemma) *Fix two positive numbers ε and δ with $\varepsilon < \delta$. For an arbitrary interval I and any function $\varphi \in \mathrm{BMO}_\varepsilon(I)$, there exists a splitting $I = I^+ \cup I^-$ such that the whole straight-line segment $[\mathfrak{b}_{I^-}(\varphi), \mathfrak{b}_{I^+}(\varphi)]$ is inside Ω_δ. Moreover, the parameters of splitting $\alpha^{\pm} \stackrel{\text{def}}{=} |I^{\pm}|/|I|$ are separated from 0 and 1 by constants depending on ε and δ only, i.e., uniformly with respect to the choice of I and φ.*

Proof Fix an interval I and a function $\varphi \in \mathrm{BMO}_\varepsilon(I)$. We now demonstrate an algorithm to find a splitting $I = I^- \cup I^+$ (i.e., choose the splitting parameters $\alpha^\pm = |I^\pm|/|I|$) so that the statement of the lemma holds. For simplicity, put $x^0 = \mathfrak{b}_I$ and $x^\pm = \mathfrak{b}_{I^\pm}$.

First, we take $\alpha^- = \alpha^+ = \frac{1}{2}$ (see Figure 1.1). If the whole segment $[x^-, x^+]$ is in Ω_δ, we fix this splitting. Assuming it is not the case, i.e., there exists a point x on this segment with $x_2 - x_1^2 > \delta^2$. Observe that only one of the segments, either $[x^-, x^0]$ or $[x^+, x^0]$, contains such points. Denote the corresponding endpoint (x^- or x^+) by ξ and define a function ρ by

$$\rho(\alpha^+) = \max_{x\in[x^-,\,x^+]}\{x_2 - x_1^2\} = \max_{x\in[\xi,\,x^0]}\{x_2 - x_1^2\}.$$

By assumption, $\rho\left(\frac{1}{2}\right) > \delta^2$.

Recall that our test function φ is fixed and the position of the Bellman points $x^\pm$ depends on the splitting parameter α^+ only. We will now change α^+ so that ξ approaches x^0, i.e., we will increase α^+ if $\xi = x^+$ and decrease it if $\xi = x^-$. We stop when $\rho(\alpha^+) = \delta^2$ and fix that splitting. It remains to check that such a moment occurs and that the corresponding α^+ is separated from 0 and 1.

Without loss of generality, assume that $\xi = x^+$. Since the function $x^+(\alpha^+)$ is continuous on the interval $(0, 1]$ and $x^+(1) = x^0$, ρ is continuous on $[\frac{1}{2}, 1]$. We have $\rho\left(\frac{1}{2}\right) > \delta^2$ and we also know that $\rho(1) \le \varepsilon^2 < \delta^2$ (because $x^0 \in \Omega_\varepsilon$). Therefore, there is a point $\alpha^+ \in \left[\frac{1}{2}, 1\right]$ with $\rho(\alpha^+) = \delta^2$ (Figure 1.2).

Having just proved that the desired point exists, we need to check that the corresponding α^+ is not too close to 0 or 1. If $\xi = x^+$, we have $\alpha^+ > \frac{1}{2}$ and $\xi_1 - x_1^0 = x_1^+ - x_1^0 = \alpha^-(x_1^+ - x_1^-)$. Similarly, if $\xi = x^-$, we have $\alpha^- > \frac{1}{2}$ and $\xi_1 - x_1^0 = x_1^- - x_1^0 = \alpha^+(x_1^- - x_1^+)$. Thus, $|\xi_1 - x_1^0| = \min\{\alpha^\pm\}|x_1^- - x_1^+|$. For the stopping value of α^+, the straight line through the points x^-, x^+, and

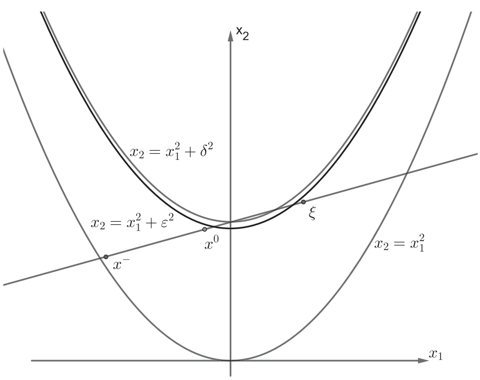


Figure 1.1 The initial splitting: $\alpha^- = \alpha^+ = \frac{1}{2}$, $\xi = x^+$.

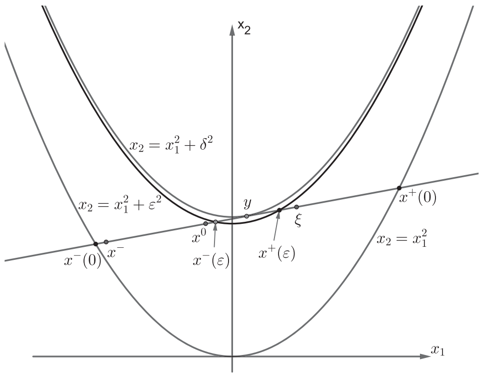


Figure 1.2 The stopping time: $[x^-, \xi]$ is tangent to the parabola $x_2 = x_1^2 + \varepsilon^2$.

x^0 is tangent to the parabola $x_2 = x_1^2 + \delta^2$ at some point y. The equation of this line is, therefore, $x_2 = 2x_1y_1 - y_1^2 + \delta^2$. The line intersects the graph of $x_2 = x_1^2 + s^2$ at the points

$$x^\pm(s) = \left(y_1 \pm \sqrt{\delta^2 - s^2},\, y_2 \pm 2y_1\sqrt{\delta^2 - s^2}\right).$$

Let us focus on the points $x^\pm(0)$ and $x^\pm(\varepsilon)$. We have

$$[x^-(\varepsilon),\, x^+(\varepsilon)] \subset [x^0,\, \xi] \subset [x^-,\, x^+] \subset [x^-(0),\, x^+(0)]$$

and, therefore,

$$\begin{aligned} 2\sqrt{\delta^2 - \varepsilon^2} &= |x_1^+(\varepsilon) - x_1^-(\varepsilon)| \le |x_1^0 - \xi_1| = \min\{\alpha^\pm\}|x_1^+ - x_1^-| \\ &\le \min\{\alpha^\pm\}|x_1^+(0) - x_1^-(0)| = \min\{\alpha^\pm\}2\delta, \end{aligned}$$

which implies

$$\sqrt{1 - \left(\frac{\varepsilon}{\delta}\right)^2} \le \alpha^+ \le 1 - \sqrt{1 - \left(\frac{\varepsilon}{\delta}\right)^2}.$$

As promised, this estimate does not depend on φ or I. □

From now on, we shall consider not the dyadic Bellman function $\mathbf{B}$, but the "true" one:

$$\mathbf{B}(x;\varepsilon) \stackrel{\text{def}}{=} \sup_{\varphi \in \mathrm{BMO}_\varepsilon(J)} \left\{\langle e^\varphi\rangle_J : \langle\varphi\rangle_J = x_1,\ \langle\varphi^2\rangle_J = x_2\right\}.$$

The test functions now run over the ε-ball of the non-dyadic BMO.

Using the splitting lemma, we are able to make the Bellman induction work in the non-dyadic case.

Lemma 1.3.7 (Bellman Induction) *If B is a continuous, locally concave function on the domain Ω_δ, satisfying the boundary condition* (1.3.3), *then* $\mathbf{B}(x;\varepsilon) \le B(x)$ *for all* $\varepsilon < \delta$.

Proof Fix a function $\varphi \in \mathrm{BMO}_\varepsilon(I)$. By the splitting lemma, we can split every subinterval $J \subset I$ in such a way that the segment $[\mathfrak{b}_{J^-}, \mathfrak{b}_{J^+}]$ is inside Ω_δ. Since B is locally concave, we have

$$|J|B(\mathfrak{b}_J) \ge |J^+|B(\mathfrak{b}_{J^+}) + |J^-|B(\mathfrak{b}_{J^-})$$

for any such splitting. Now we can repeat, word for word, the arguments used in the dyadic case. Recall that $\mathcal{I}_n$ stands for the set of intervals of n-th generation, then

$$\begin{aligned} |I|B(\mathfrak{b}_I) &\ge |I^+|B(\mathfrak{b}_{I^+}) + |I^-|B(\mathfrak{b}_{I^-}) \\ &\ge \sum_{J \in \mathcal{I}_n} |J|B(\mathfrak{b}_J) = \int_I B(x^{(n)}(s))\, ds, \end{aligned}$$

where $x^{(n)}(s) = \mathfrak{b}_J$, when $s \in J$, $J \in \mathcal{I}_n$. By the Lebesgue differentiation theorem, we have $x^{(n)}(s) \to (\varphi(s), \varphi^2(s))$ almost everywhere. (We have used here the fact that we split the intervals so that all coefficients $\alpha^\pm$ are uniformly separated from 0 and 1, and, therefore, $\max\{|J|\colon J \in \mathcal{I}_n\} \to 0$ as $n \to \infty$.) Now, we can pass to the limit in this inequality as $n \to \infty$. Again, first we assume φ to be bounded and, by the Lebesgue dominated convergence theorem, pass to the limit in the integral using the boundary condition (1.3.3)

$$|I|B(\mathfrak{b}_I(\varphi)) \ge \int_I B(\varphi(s), \varphi^2(s))\, ds = \int_I e^{\varphi(s)} ds = |I|\langle e^\varphi \rangle_I.$$

Then using the cut-off approximation, we get the same inequality for an arbitrary $\varphi \in \mathrm{BMO}_\varepsilon(I)$ such that $\mathfrak{b}_I(\varphi) = x$ for any given $x \in \Omega_\varepsilon$. □

Corollary 1.3.8

$$\mathbf{B}(x;\varepsilon) \le B(x;\delta), \qquad \varepsilon < \delta < 1.$$

Proof The function $B(x;\delta)$ was constructed as a locally concave function satisfying boundary condition (1.3.3). □

Corollary 1.3.9

$$\mathbf{B}(x;\varepsilon) \le B(x;\varepsilon). \tag{1.3.11}$$

Proof Since the function $B(x;\delta)$ is continuous with respect to the parameter $\delta \in (0,1)$, we can pass to the limit $\delta \to \varepsilon$ in the preceding corollary. □

Now, we would like to prove the inequality converse to (1.3.11). To this end, for every point x of Ω_ε, we construct a test function φ with BMO-norm ε, satisfying $\langle e^\varphi \rangle = B(x;\varepsilon)$, and such that its Bellman point is x. This would imply the inequality $\mathbf{B}(x;\varepsilon) \geq B(x;\varepsilon)$. Recall that such a test function that realizes the extremal value for the functional under investigation is called an *extremizer*.

First, we construct an extremizer φ_0 for the point $(0,\varepsilon^2)$. Without loss of generality, we can work on $I = [0,1]$. Note that the function $\varphi_a \stackrel{\text{def}}{=} \varphi_0 + a$ will then be an extremizer for the point $(a, a^2+\varepsilon^2)$. Indeed, φ_a has the same norm as φ_0, and if

$$\langle e^{\varphi_0} \rangle = B(0,\varepsilon^2;\varepsilon) = \frac{e^{-\varepsilon}}{1-\varepsilon},$$

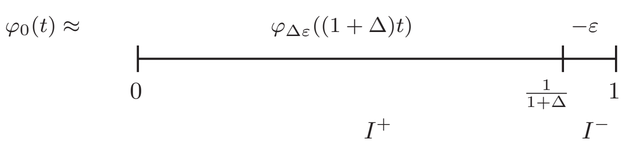
then

$$\langle e^{\varphi_a} \rangle = \frac{e^{a-\varepsilon}}{1-\varepsilon} = B(a, a^2+\varepsilon^2;\varepsilon).$$

The point $(0,\varepsilon^2)$ is on the extremal line starting at $(-\varepsilon,\varepsilon^2)$. To keep equality on each step of the Bellman induction, when we split I into two subintervals I^- and I^+, the segment $[x^-,\, x^+]$ has to be contained in the extremal line along which our function B is linear. Since x is a convex combination of x^- and x^+, one of these points, say x^+, has to be to the right of x. However, the extremal line ends at $x = (0,\varepsilon^2)$, and so there seems to be nowhere to place that point. We circumvent this difficulty by placing x^+ infinitesimally close to x and using an approximation procedure. Where should x^- be placed? We already know extremizers for points on the lower boundary $x_2 = x_1^2$, since the only test function there are constants. Thus, it is convenient to put x^- there. Therefore, we set

$$x^- = (-\varepsilon,\varepsilon^2) \qquad \text{and} \qquad x^+ = (\Delta\varepsilon,\varepsilon^2),$$

for small Δ. To get these two points, we have to split I in proportion $1:\Delta$, that is, we take $I^+ = [0, \frac{1}{1+\Delta}]$ and $I^- = [\frac{1}{1+\Delta}, 1]$. To get the point x^-, we have



to put $\varphi_0(t) = -\varepsilon$ on I^-. On I^+, we put a function corresponding not to the point x^+, but to the point $(\Delta\varepsilon, (1+\Delta^2)\varepsilon^2)$ on the upper boundary, which is close to x^+ (the distance between these two points is of order Δ^2). For such a point, the extremal function is $\varphi_{\Delta\varepsilon}(t) = \varphi_0(t) + \Delta\varepsilon$. Therefore, this function, when properly rescaled, can be placed on I^+. As a result, we obtain

$$\varphi_0(t) \approx \varphi_0\big((1+\Delta)t\big) + \Delta\varepsilon \approx \varphi_0(t) + \varphi_0'(t)\Delta t + \Delta\varepsilon,$$

which yields

$$\varphi_0'(t) = -\frac{\varepsilon}{t}.$$

Taking into account the boundary condition $\varphi_0(1) = -\varepsilon$, we get

$$\varphi_0(t) = \varepsilon \log \frac{1}{t} - \varepsilon.$$

Let us check whether we have found what we need. By the direct calculation, we get: $\langle \varphi_0 \rangle_{[0,1]} = 0$, $\langle \varphi_0^2 \rangle_{[0,1]} = \varepsilon^2$, and

$$\langle e^{\varphi_0} \rangle_{[0,1]} = \int_0^1 e^{-\varepsilon} \frac{dt}{t^\varepsilon} = \frac{e^{-\varepsilon}}{1-\varepsilon} = B(0, \varepsilon^2; \varepsilon).$$

It is easy now to get an extremal function for an arbitrary point x in Ω_ε. First of all, we draw the extremal line through x. It touches the upper boundary at the point $(a, a^2 + \varepsilon^2)$ with $a = x_1 + \sqrt{\varepsilon^2 - x_2 + x_1^2}$ and intersects the lower boundary at the point (u, u^2) with $u = a - \varepsilon$. Now, we split the interval $[0,1]$ in the ratio $(x_1 - u) : (a - x_1)$ and concatenate the two known extremizers, $\varphi = u$ for the $x^- = (u, u^2)$ and $\varphi = \varphi_a$ for $x^+ = (a, a^2 + \varepsilon^2)$. This gives the following function:

$$\varphi(t) = \begin{cases} \varepsilon \log \frac{x_1 - u}{\varepsilon t} + u & \text{for } 0 \le t \le \frac{x_1 - u}{\varepsilon} \\ u & \text{for } \frac{x_1 - u}{\varepsilon} \le t \le 1 \end{cases}, \tag{1.3.12}$$

where

$$u = x_1 + \sqrt{\varepsilon^2 - x_2 + x_1^2} - \varepsilon.$$

This is a function from BMO_ε satisfying the required property $\langle e^\varphi \rangle_{[0,1]} = B(x; \varepsilon)$ (see Problem 1.3.4 later on).

This completes the proof of the following theorem.

THEOREM 1.3.10 *If* $\varepsilon < 1$, *then*

$$\mathbf{B}(x; \varepsilon) = \frac{1 - \sqrt{\varepsilon^2 - x_2 + x_1^2}}{1 - \varepsilon} \exp\left\{ x_1 + \sqrt{\varepsilon^2 - x_2 + x_1^2} - \varepsilon \right\};$$

if $\varepsilon \ge 1$, *then* $\mathbf{B}(x; \varepsilon) = \infty$.

Indeed, the second statement can be verified by the same extremal function φ because e^φ is not summable on $[0,1]$ for $\varepsilon \ge 1$.

Historical Remarks

The first proof of the theorem mentioned earlier was independently presented in [168] and [183]; a complete proof of this result together with the estimate from below (i.e., the lower Bellman function) and consideration of the dyadic version of the problem can be found in [171]. The Bellman function for the classical weak form of the John–Nirenberg inequality can be found in [185] or in [191].

Exercises

PROBLEM 1.3.1 Calculate the quadratic form of the Hessian $\sum_{i,j} B_{x_i x_j}\Delta_i\Delta_j$ for the function $B = B_+(x,\delta)$.

PROBLEM 1.3.2 Find the extremal trajectories along which the Hessian degenerates. Check that these are the tangent line to the parabola $y_2 = y_1^2 + \delta^2$.

PROBLEM 1.3.3 Check that for the function $B = B_+(x;\delta)$ from (1.3.10), the main inequality (1.3.1) is not true for some points. In particular,

$$B(u, u^2+\delta^2;\delta) \leq \frac{B(u-\frac{\delta}{\sqrt{2}},(u-\frac{\delta}{\sqrt{2}})^2;\delta)+B(u+\frac{\delta}{\sqrt{2}},(u+\frac{\delta}{\sqrt{2}})^2+\delta^2;\delta)}{2}.$$

PROBLEM 1.3.4 Verify the following properties of the extremal function φ :

- $\langle\varphi\rangle_{[0,1]} = x_1$;
- $\langle\varphi^2\rangle_{[0,1]} = x_2$;
- $\langle e^\varphi\rangle_{[0,1]} = B(x_1, x_2;\varepsilon)$;
- $\varphi \in \mathrm{BMO}_\varepsilon$.

PROBLEM 1.3.5 Recall that we also obtained a second solution in (1.3.10):

$$B_-(x;\varepsilon) = \frac{1+\sqrt{\varepsilon^2 - x_2 + x_1^2}}{1+\varepsilon}\,\exp\Big\{x_1 - \sqrt{\varepsilon^2 - x_2 + x_1^2} + \varepsilon\Big\}.$$

Check that this is the solution of the following extremal problem:

$$\mathbf{B}_{\min}(x;\varepsilon) \stackrel{\text{def}}{=} \inf_{\varphi\in\mathrm{BMO}_\varepsilon(I)} \left\{\langle e^\varphi\rangle_I : \langle\varphi\rangle_I = x_1,\ \langle\varphi^2\rangle_I = x_2\right\},$$

that is, check that the Bellman induction works and construct an extremal function for every $x \in \Omega_\varepsilon$.