国际知名大学原版教材

——信息技术学科与电气工程学科系列

Abstract Dynamic Programming

抽象动态规划

Dimitri P. Bertsekas 著

清华大学出版社

国际知名大学原版教材
——信息技术学科与电气工程学科系列

Abstract Dynamic Programming

抽象动态规划

Dimitri P. Bertsekas 著

清华大学出版社
北京

北京市版权局著作权合同登记号:01-2014-2209

图书在版编目(CIP)数据

抽象动态规划 = Abstract dynamic programming: 英文/(美)博塞克斯(Bertsekas, D. P.)著. —北京:清华大学出版社. 2014

国际知名大学原版教材·信息技术学科与电气工程学科系列

ISBN 978-7-302-36269-2

Ⅰ. ①抽… Ⅱ. ①博… Ⅲ. ①动态规划—高等学校—教材—英文 Ⅳ. ①O221.3

中国版本图书馆 CIP 数据核字(2014)第 077153 号

责任编辑:王一玲
封面设计:傅瑞学
责任校对:白 蕾
责任印制:沈 露

出版发行:清华大学出版社
网 址:http://www.tup.com.cn, http://www.wqbook.com
地 址:北京清华大学学研大厦 A 座 **邮 编**:100084
社总机:010-62770175 **邮 购**:010-62786544
投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn
质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn
印 装 者:北京嘉实印刷有限公司
经 销:全国新华书店
开 本:153mm×235mm **印 张**:16.25 **字 数**:272 千字
版 次:2014 年 7 月第 1 版 **印 次**:2014 年 7 月第1次印刷
印 数:1～2000
定 价:39.00 元

产品编号:057276-01

Abstract Dynamic Programming

影印版序

本书采用一种简洁的方式介绍动态规划的理论和方法。作者首先把动态规划的核心问题表述为一类抽象映射的不动点问题；然后将决定不动点问题求解难度的主要因素概括为上述抽象映射的两个性质：单调性和压缩性；接着在假设单调性始终成立的前提下，围绕压缩性是否成立，顺序讨论了各种典型情况下相应不动点问题的主要性质和求解方法。其中第 2 章介绍压缩性成立时的结果，第 3 章介绍压缩性部分成立时的结果，第 4 章介绍压缩性不成立时的结果，最后在第 5 章介绍了策略受限情况的一些结果。这些内容涉及不动点的存在性、值迭代方法和策略迭代方法的收敛性以及多种常用近似方法的误差上界等动态规划的基本问题。

本书作者是美国麻省理工学院电气工程和计算机科学系的资深教授，在线性规划、非线性规划、动态规划、网络优化、凸分析与优化等众多优化领域著有十余部专著或教科书。如同作者其他著作一样，本书在描述问题、定义概念和证明定理时力求清晰、严谨和完整。尽管本书始终以不动点问题为讨论对象，但每部分内容都给出了相应的动态规划实例。结合这些例子，很容易理解所获得的结果和动态规划问题的关系。因此，对于具有一定数学基础的读者，既可以把本书作为深入了解动态规划理论的专著，也可以将其作为自学动态规划知识的教材。

动态规划是解决复杂优化问题的一种基本方法。同线性规划、非线性规划、网络优化等其他优化领域的基本理论相比，应用动态规划方法解决优化问题的原理相对而言比较简单。但对同样的问题，采用不同的建模和求解策略，所产生的实际效果可能存在很大差异。因此，采用动态规划方法解决具体问题时具有很大的灵活性。通过阅读本书，系统掌握动态规划的核心理论和方法，对于更好地应用动态规划思想和方法解决实际问题，一定大有裨益。

清华大学自动化系　王书宁教授

2014 年 3 月 15 日

Preface

This book aims at a unified and economical development of the core theory and algorithms of total cost sequential decision problems, based on the strong connections of the subject with fixed point theory. The analysis focuses on the abstract mapping that underlies dynamic programming (DP for short) and defines the mathematical character of the associated problem. Our discussion centers on two fundamental properties that this mapping may have: *monotonicity* and (weighted sup-norm) *contraction*. It turns out that the nature of the analytical and algorithmic DP theory is determined primarily by the presence or absence of these two properties, and the rest of the problem's structure is largely inconsequential.

In this book, with some minor exceptions, we will assume that monotonicity holds. Consequently, we organize our treatment around the contraction property, and we focus on four main classes of models:

(a) **Contractive models**, discussed in Chapter 2, which have the richest and strongest theory, and are the benchmark against which the theory of other models is compared. Prominent among these models are discounted stochastic optimal control problems. The development of these models is quite thorough and includes the analysis of recent approximation algorithms for large-scale problems (neuro-dynamic programming, reinforcement learning).

(b) **Semicontractive models**, discussed in Chapter 3 and parts of Chapter 4. The term "semicontractive" is used qualitatively here, to refer to a variety of models where some policies have a regularity/contraction-like property but others do not. A prominent example is stochastic shortest path problems, where one aims to drive the state of a Markov chain to a termination state at minimum expected cost. These models also have a strong theory under certain conditions, often nearly as strong as those of the contractive models.

(c) **Noncontractive models**, discussed in Chapter 4, which rely on just monotonicity. These models are more complex than the preceding ones and much of the theory of the contractive models generalizes in weaker form, if at all. For example, in general the associated Bellman equation need not have a unique solution, the value iteration method may work starting with some functions but not with others, and the policy iteration method may not work at all. Infinite horizon examples of these models are the classical positive and negative DP problems, first analyzed by Dubins and Savage, Blackwell, and

Strauch, which are discussed in various sources. Some new semicontractive models are also discussed in this chapter, further bridging the gap between contractive and noncontractive models.

(d) **Restricted policies and Borel space models**, which are discussed in Chapter 5. These models are motivated in part by the complex measurability questions that arise in mathematically rigorous theories of stochastic optimal control involving continuous probability spaces. Within this context, the admissible policies and DP mapping are restricted to have certain measurability properties, and the analysis of the preceding chapters requires modifications. Restricted policy models are also useful when there is a special class of policies with favorable structure, which is "closed" with respect to the standard DP operations, in the sense that analysis and algorithms can be confined within this class.

We do not consider average cost DP problems, whose character bears a much closer connection to stochastic processes than to total cost problems. We also do not address specific stochastic characteristics underlying the problem, such as for example a Markovian structure. Thus our results apply equally well to Markovian decision problems and to sequential minimax problems. While this makes our development general and a convenient starting point for the further analysis of a variety of different types of problems, it also ignores some of the interesting characteristics of special types of DP problems that require an intricate probabilistic analysis.

Let us describe the research content of the book in summary, deferring a more detailed discussion to the end-of-chapter notes. A large portion of our analysis has been known for a long time, but in a somewhat fragmentary form. In particular, the contractive theory, first developed by Denardo [Den67], has been known for the case of the unweighted sup-norm, but does not cover the important special case of stochastic shortest path problems where all policies are proper. Chapter 2 transcribes this theory to the weighted sup-norm contraction case. Moreover, Chapter 2 develops extensions of the theory to approximate DP, and includes material on asynchronous value iteration (based on the author's work [Ber82], [Ber83]), and asynchronous policy iteration algorithms (based on the author's joint work with Huizhen (Janey) Yu [BeY10a], [BeY10b], [YuB11a]). Most of this material is relatively new, having been presented in the author's recent book [Ber12a] and survey paper [Ber12b], with detailed references given there. The analysis of infinite horizon noncontractive models in Chapter 4 was first given in the author's paper [Ber77], and was also presented in the book by Bertsekas and Shreve [BeS78], which in addition contains much of the material on finite horizon problems, restricted policies models, and Borel space models. These were the starting point and main sources for our development.

The new research presented in this book is primarily on the semi-

contractive models of Chapter 3 and parts of Chapter 4. Traditionally, the theory of total cost infinite horizon DP has been bordered by two extremes: discounted models, which have a contractive nature, and positive and negative models, which do not have a contractive nature, but rely on an enhanced monotonicity structure (monotone increase and monotone decrease models, or in classical DP terms, positive and negative models). Between these two extremes lies a gray area of problems that are not contractive, and either do not fit into the categories of positive and negative models, or possess additional structure that is not exploited by the theory of these models. Included are stochastic shortest path problems, search problems, linear-quadratic problems, a host of queueing problems, multiplicative and exponential cost models, and others. Together these problems represent an important part of the infinite horizon total cost DP landscape. They possess important theoretical characteristics, not generally available for positive and negative models, such as the uniqueness of solution of Bellman's equation within a subset of interest, and the validity of useful forms of value and policy iteration algorithms.

Our semicontractive models aim to provide a unifying abstract DP structure for problems in this gray area between contractive and noncontractive models. The analysis is motivated in part by stochastic shortest path problems, where there are two types of policies: *proper*, which are the ones that lead to the termination state with probability one from all starting states, and *improper*, which are the ones that are not proper. Proper and improper policies can also be characterized through their Bellman equation mapping: for the former this mapping is a contraction, while for the latter it is not. In our more general semicontractive models, policies are also characterized in terms of their Bellman equation mapping, through a notion of *regularity*, which generalizes the notion of a proper policy and is related to classical notions of asymptotic stability from control theory.

In our development a policy is regular within a certain set if its cost function is the unique asymptotically stable equilibrium (fixed point) of the associated DP mapping within that set. *We assume that some policies are regular while others are not*, and impose various assumptions to ensure that attention can be focused on the regular policies. From an analytical point of view, this brings to bear the theory of fixed points of monotone mappings. From the practical point of view, this allows application to a diverse collection of interesting problems, ranging from stochastic shortest path problems of various kinds, where the regular policies include the proper policies, to linear-quadratic problems, where the regular policies include the stabilizing linear feedback controllers.

The definition of regularity is introduced in Chapter 3, and its theoretical ramifications are explored through extensions of the classical stochastic shortest path and search problems. In Chapter 4, semicontractive models are discussed in the presence of additional monotonicity structure, which brings to bear the properties of positive and negative DP models. With the

aid of this structure, the theory of semicontractive models can be strengthened and can be applied to several additional problems, including risk-sensitive/exponential cost problems.

The book has a theoretical research monograph character, but requires a modest mathematical background for all chapters except the last one, essentially a first course in analysis. Of course, prior exposure to DP will definitely be very helpful to provide orientation and context. A few exercises have been included, either to illustrate the theory with examples and counterexamples, or to provide applications and extensions of the theory. Solutions of all the exercises can be found in Appendix D, at the book's internet site

http://www.athenasc.com/abstractdp.html

and at the author's web site

http://web.mit.edu/dimitrib/www/home.html

Additional exercises and other related material may be added to these sites over time.

I would like to express my appreciation to a few colleagues for interactions, recent and old, which have helped shape the form of the book. My collaboration with Steven Shreve on our 1978 book provided the motivation and the background for the material on models with restricted policies and associated measurability questions. My collaboration with John Tsitsiklis on stochastic shortest path problems provided inspiration for the work on semicontractive models. My collaboration with Janey Yu played an important role in the book's development, and is reflected in our joint work on asynchronous policy iteration, on perturbation models, and on risk-sensitive models. Moreover Janey contributed significantly to the material on semicontractive models with many insightful suggestions. Finally, I am thankful to Mengdi Wang, who went through portions of the book with care, and gave several helpful comments.

Dimitri P. Bertsekas, Spring 2013

NOTE ADDED TO THE CHINESE EDITION

The errata of the original edition, as per March 1, 2014, have been incorporated in the present edition of the book. The following two papers have a strong connection to the book, and amplify on the range of applications of the semicontractive models of Chapters 3 and 4:

(1) D. P. Bertsekas, "Robust Shortest Path Planning and Semicontractive Dynamic Programming," Lab. for Information and Decision Systems Report LIDS-P-2915, MIT, Feb. 2014.

(2) D. P. Bertsekas, "Infinite-Space Shortest Path Problems and Semicontractive Dynamic Programming," Lab. for Information and Decision Systems Report LIDS-P-2916, MIT, Feb. 2014.

These papers may be viewed as "on-line appendixes" of the book. They can be downloaded from the book's internet site and the author's web page.

Contents

1

Introduction

Contents

1.1 STRUCTURE OF DYNAMIC PROGRAMMING PROBLEMS

Dynamic programming (DP for short) is the principal method for analysis of a large and diverse class of sequential decision problems. Examples are deterministic and stochastic optimal control problems with a continuous state space, Markov and semi-Markov decision problems with a discrete state space, minimax problems, and sequential zero sum games. While the nature of these problems may vary widely, their underlying structures turn out to be very similar. In all cases there is an underlying mapping that depends on an associated controlled dynamic system and corresponding cost per stage. This mapping, the DP operator, provides a "compact signature" of the problem. It defines the cost function of policies and the optimal cost function, and it provides a convenient shorthand notation for algorithmic description and analysis.

More importantly, the structure of the DP operator defines the mathematical character of the associated problem. The purpose of this book is to provide an analysis of this structure, centering on two fundamental properties: *monotonicity* and (weighted sup-norm) *contraction*. It turns out that the nature of the analytical and algorithmic DP theory is determined primarily by the presence or absence of these two properties, and the rest of the problem's structure is largely inconsequential.

A Deterministic Optimal Control Example

To illustrate our viewpoint, let us consider a discrete-time deterministic optimal control problem described by a system equation

$$x_{k+1} = f(x_k, u_k), \qquad k = 0, 1, \ldots \tag{1.1}$$

Here x_k is the state of the system taking values in a set X (the state space), and u_k is the control taking values in a set U (the control space). At stage k, there is a cost

$$\alpha^k g(x_k, u_k)$$

incurred when u_k is applied at state x_k, where α is a scalar in $(0, 1]$ that has the interpretation of a discount factor when $\alpha < 1$. The controls are chosen as a function of the current state, subject to a constraint that depends on that state. In particular, at state x the control is constrained to take values in a given set $U(x) \subset U$. Thus we are interested in optimization over the set of (nonstationary) policies

$$\Pi = \{\{\mu_0, \mu_1, \ldots\} \mid \mu_k \in \mathcal{M},\ k = 0, 1, \ldots\},$$

where $\mathcal{M}$ is the set of functions $\mu : X \mapsto U$ defined by

$$\mathcal{M} = \{\mu \mid \mu(x) \in U(x),\ \forall\ x \in X\}.$$

The total cost of a policy $\pi = \{\mu_0, \mu_1, \ldots\}$ over an infinite number of stages and starting at an initial state x_0 is

$$J_\pi(x_0) = \sum_{k=0}^{\infty} \alpha^k g\big(x_k, \mu_k(x_k)\big), \tag{1.2}$$

where the state sequence $\{x_k\}$ is generated by the deterministic system (1.1) under the policy π:

$$x_{k+1} = f\big(x_k, \mu_k(x_k)\big), \quad k = 0, 1, \ldots$$

The optimal cost function is †

$$J^*(x) = \inf_{\pi \in \Pi} J_\pi(x), \qquad x \in X.$$

For any policy $\pi = \{\mu_0, \mu_1, \ldots\}$, consider the policy $\pi_1 = \{\mu_1, \mu_2, \ldots\}$ and write by using Eq. (1.2),

$$J_\pi(x) = g\big(x, \mu_0(x)\big) + \alpha J_{\pi_1}\big(f(x, \mu_0(x)\big).$$

We have for all $x \in X$

$$\begin{aligned} J^*(x) &= \inf_{\pi = \{\mu_0, \pi_1\} \in \Pi} \Big\{ g\big(x, \mu_0(x)\big) + \alpha J_{\pi_1}\big(f(x, \mu_0(x)\big) \Big\} \\ &= \inf_{\mu_0 \in \mathcal{M}} \Big\{ g\big(x, \mu_0(x)\big) + \alpha \inf_{\pi_1 \in \Pi} J_{\pi_1}\big(f(x, \mu_0(x)\big) \Big\} \\ &= \inf_{\mu_0 \in \mathcal{M}} \Big\{ g\big(x, \mu_0(x)\big) + \alpha J^*\big(f(x, \mu_0(x)\big) \Big\}. \end{aligned}$$

The minimization over $\mu_0 \in \mathcal{M}$ can be written as minimization over all $u \in U(x)$, so we can write the preceding equation as

$$J^*(x) = \inf_{u \in U(x)} \Big\{ g(x, u) + \alpha J^*\big(f(x, u)\big) \Big\}, \qquad \forall\, x \in X. \tag{1.3}$$

This equation is an example of *Bellman's equation*, which plays a central role in DP analysis and algorithms. If it can be solved for J^*, an optimal stationary policy $\{\mu^*, \mu^*, \ldots\}$ may typically be obtained by minimization of the right-hand side for each x, i.e.,

$$\mu^*(x) \in \arg\min_{u \in U(x)} \Big\{ g(x, u) + \alpha J^*\big(f(x, u)\big) \Big\}, \qquad \forall\, x \in X. \tag{1.4}$$

† For the informal discussion of this section, we will disregard a few mathematical issues. In particular, we assume that the series defining J_π in Eq. (1.2) is convergent for all allowable π, and that the optimal cost function J^* is real-valued. We will address such issues later.

We now note that both Eqs. (1.3) and (1.4) can be stated in terms of the expression

$$H(x,u,J) = g(x,u) + \alpha J\big(f(x,u)\big), \qquad x \in X,\ u \in U(x).$$

Defining

$$(T_\mu J)(x) = H\big(x,\mu(x),J\big), \qquad x \in X,$$

and

$$(TJ)(x) = \inf_{u\in U(x)} H(x,u,J) = \inf_{\mu\in\mathcal{M}} (T_\mu J)(x), \qquad x \in X,$$

we see that Bellman's equation (1.3) can be written compactly as

$$J^* = TJ^*,$$

i.e., J^* is the fixed point of T, viewed as a mapping from the set of real-valued functions on X into itself. Moreover, it can be similarly seen that J_μ, the cost function of the stationary policy $\{\mu,\mu,\ldots\}$, is a fixed point of T_μ. In addition, the optimality condition (1.4) can be stated compactly as

$$T_{\mu^*} J^* = TJ^*.$$

We will see later that additional properties, as well as a variety of algorithms for finding J^* can be analyzed using the mappings T and T_μ.

One more property that holds in some generality is worth noting. For a given policy $\pi = \{\mu_0,\mu_1,\ldots\}$ and a terminal cost $\alpha^N \bar{J}(x_N)$ for the state x_N at the end of N stages, consider the N-stage cost function

$$J_{\pi,N}(x_0) = \alpha^N \bar{J}(x_N) + \sum_{k=0}^{N-1} \alpha^k g\big(x_k,\mu_k(x_k)\big). \tag{1.5}$$

Then it can be verified by induction that for all initial states x_0, we have

$$J_{\pi,N}(x_0) = (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} \bar{J})(x_0). \tag{1.6}$$

Here $T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}}$ is the composition of the mappings $T_{\mu_0}, T_{\mu_1}, \ldots T_{\mu_{N-1}}$, i.e., for all J,

$$(T_{\mu_0} T_{\mu_1} J)(x) = \big(T_{\mu_0}(T_{\mu_1} J)\big)(x), \qquad x \in X,$$

and more generally

$$(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} J)(x) = \big(T_{\mu_0}(T_{\mu_1}(\cdots(T_{\mu_{N-1}} J)))\big)(x), \qquad x \in X,$$

(our notational conventions are summarized in Appendix A). Thus the finite horizon cost functions $J_{\pi,N}$ of π can be defined in terms of the mappings T_μ [cf. Eq. (1.6)], and so can their infinite horizon limit J_π:

$$J_\pi(x) = \lim_{N\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} \bar{J})(x), \qquad x \in X, \tag{1.7}$$

where $\bar{J}$ is the zero function, $\bar{J}(x) = 0$ for all $x \in X$ (assuming the limit exists).

Connection with Fixed Point Methodology

The Bellman equation (1.3) and the optimality condition (1.4), stated in terms of the mappings T_μ and T, highlight the central theme of this book, which is that DP theory is intimately connected with the theory of abstract mappings and their fixed points. Analogs of the Bellman equation, $J^* = TJ^*$, optimality conditions, and other results and computational methods hold for a great variety of DP models, and can be stated compactly as described above in terms of the corresponding mappings T_μ and T. The gain from this abstraction is greater generality and mathematical insight, as well as a more unified, economical, and streamlined analysis.

1.2 ABSTRACT DYNAMIC PROGRAMMING MODELS

In this section we formally introduce and illustrate with examples an abstract DP model, which embodies the ideas discussed in the preceding section.

1.2.1 Problem Formulation

Let X and U be two sets, which we loosely refer to as a set of "states" and a set of "controls," respectively. For each $x \in X$, let $U(x) \subset U$ be a nonempty subset of controls that are feasible at state x. We denote by $\mathcal{M}$ the set of all functions $\mu : X \mapsto U$ with $\mu(x) \in U(x)$, for all $x \in X$.

In analogy with DP, we refer to sequences $\pi = \{\mu_0, \mu_1, \ldots\}$, with $\mu_k \in \mathcal{M}$ for all k, as "nonstationary policies," and we refer to a sequence $\{\mu, \mu, \ldots\}$, with $\mu \in \mathcal{M}$, as a "stationary policy." In our development, stationary policies will play a dominant role, and with slight abuse of terminology, we will also refer to any $\mu \in \mathcal{M}$ as a "policy" when confusion cannot arise.

Let $R(X)$ be the set of real-valued functions $J : X \mapsto \Re$, and let $H : X \times U \times R(X) \mapsto \Re$ be a given mapping.† For each policy $\mu \in \mathcal{M}$, we consider the mapping $T_\mu : R(X) \mapsto R(X)$ defined by

$$(T_\mu J)(x) = H\big(x, \mu(x), J\big), \qquad \forall\, x \in X,\, J \in R(X),$$

and we also consider the mapping T defined by‡

$$(TJ)(x) = \inf_{u \in U(x)} H(x, u, J), \qquad \forall\, x \in X,\, J \in R(X).$$

---

† Our notation and mathematical conventions are outlined in Appendix A. In particular, we denote by $\Re$ the set of real numbers, and by $\Re^n$ the space of n-dimensional vectors with real components.

‡ We assume that H, $T_\mu J$, and TJ are real-valued for $J \in R(X)$ in the present chapter and in Chapter 2. In Chapters 3-5 we will allow $H(x, u, J)$, and hence also $(T_\mu J)(x)$ and $(TJ)(x)$, to take the values ∞ and $-\infty$.

Similar to the deterministic optimal control problem of the preceding section, the mappings T_μ and T serve to define a multistage optimization problem and a DP-like methodology for its solution. In particular, for some function $\bar{J} \in R(X)$, and nonstationary policy $\pi = \{\mu_0, \mu_1, \ldots\}$, we define for each integer $N \geq 1$ the functions

$$J_{\pi,N}(x) = (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} \bar{J})(x), \qquad x \in X,$$

where $T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}}$ denotes the composition of the mappings T_{μ_0}, T_{μ_1}, $\ldots$, $T_{\mu_{N-1}}$, i.e.,

$$T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} J = \big(T_{\mu_0}(T_{\mu_1}(\cdots(T_{\mu_{N-2}}(T_{\mu_{N-1}} J)))\cdots)\big), \quad J \in R(X).$$

We view $J_{\pi,N}$ as the "N-stage cost function" of π [cf. Eq. (1.5)]. Consider also the function

$$J_\pi(x) = \limsup_{N\to\infty} J_{\pi,N}(x) = \limsup_{N\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}} \bar{J})(x), \qquad x \in X,$$

which we view as the "infinite horizon cost function" of π [cf. Eq. (1.7); we use lim sup for generality, since we are not assured that the limit exists]. We want to minimize J_π over π, i.e., to find

$$J^*(x) = \inf_\pi J_\pi(x), \qquad x \in X,$$

and a policy π^* that attains the infimum, if one exists.

The key connection with fixed point methodology is that J^* "typically" (under mild assumptions) can be shown to satisfy

$$J^*(x) = \inf_{u\in U(x)} H(x, u, J^*), \qquad \forall\, x \in X,$$

i.e., it is a fixed point of T. We refer to this as *Bellman's equation* [cf. Eq. (1.3)]. Another fact is that if an optimal policy π^* exists, it "typically" can be selected to be stationary, $\pi^* = \{\mu^*, \mu^*, \ldots\}$, with $\mu^* \in \mathcal{M}$ satisfying an optimality condition, such as for example

$$T_{\mu^*} J^* = T J^*$$

[cf. Eq. (1.4)]. Several other results of an analytical or algorithmic nature also hold under appropriate conditions, which will be discussed in detail later.

However, Bellman's equation and other related results may not hold without T_μ and T having some special structural properties. Prominent among these are a monotonicity assumption that typically holds in DP problems, and a contraction assumption that holds for some important classes of problems.