7 MIXED CRITERIA

Eugene A. Feinberg
Adam Shwartz

Abstract: Mixed criteria are linear combinations of standard criteria which cannot be represented as standard criteria. Linear combinations of total discounted and average rewards as well as linear combinations of total discounted rewards with different discount factors are examples of mixed criteria. We discuss the structure of optimal policies and algorithms for their computation for problems with and without constraints.

7.1 INTRODUCTION

The discounted cost criterion is widely and successfully used in various application areas. When modeling economic phenomena, the discount factor models the value of money in time and is determined by return rate (or interest rate) or, in a more general context, by the "opportunity costs" which presume that a dollar now is worth more that a dollar in a year. Being invested, current funds will bring an additional return in a year. In other areas, such as the control of communications networks, the discounted criterion may reflect the imprecise but fundamental principle that future occurrences are less important than immediate ones. In reliability, discounting models systems with geometric life time distributions.

Obviously, if some part of the cost decreases (in time) at an exponential rate, then a discounted cost arises. This is the case, for example, in production processes. When a new item is manufactured, we expect some production costs to decrease as production methods are improved. Obviously there is a learning curve for all involved, along which various costs decrease. If the effect of this learning diminishes geometrically (or exponentially), then the total cost (over the infinite time-horizon) is of the discounted type. This would also be the case if the cost of obtaining a component decreases at an exponential rate. Such an exponential decrease is evident in the computers industry (and one aspect goes by "Moore's law"): prices of various components decrease at an exponential rate, and since the processing speed increases at an exponential rate, the "unit cost" of processing power decreases exponentially fast.

However, the rates (or discount factors) for these different mechanisms are clearly unrelated. When combining several such costs (such as processing speed with economic considerations), we are naturally led to deal with several different discount factors, each applicable to a component of our optimization problem. Multiple discount factors also arise in control problems for systems with multiple parallel unreliable components. For details on applications see Feinberg and Shwartz [14, 15, 17]. Similarly, in stochastic games it is natural to consider situations where each player has its own discount factor.

In contrast to the discounted criterion, the average cost measures long-term behavior and, usually is insensitive to present and short-term conditions. Naturally, this criterion is appropriate for other applications, such as the long-term performance of systems. As before, if our criteria include system performance (measured through the average cost) as well as rewards (measured through a discounted cost), we are led to problems with mixed criteria.

In this paper we review results concerning such criteria, and point out some open questions. We shall mention specific, as well as general areas of potential applications. The main theoretical questions are:

- Existence of good (optimal, ε-optimal) policies for optimization, multi-objective optimization and in particular constrained optimization problems,
- Structure of "good policies," and
- Computational schemes for the calculation of the value and good policies.

We emphasize the mixed-discounted problem, where the criteria are all of the discounted type, but with several discount factors, since it possesses a rich structure and, in addition, much is already known. In Section 7.7 we shall review some other related results. We conclude this section with a brief survey of different mixed criterion problems.

There are several treatments of discounted models where the discounting is more general than the standard one. Hinderer [27] investigates a general model where the discounting is a function of the complete history of the model. In a number of papers, see e.g. Schäl [41], Chitashvili [7], Stidham [43], or Haurie and L'Ecuyer [22], the discount rate depends on the current state and action. This type of discounting arises when a discounted semi-Markov Decision process is converted into an MDP; see Puterman [37] or Feinberg [13] for details.

Mixed criteria and, in particular, mixed discounted criteria are linear combinations of standard criteria. The first two papers dealing with mixed criteria were published in 1982. Golabi, Kulkarni, and Way [21] considered a mixture of average reward and total discounted criteria to manage a statewide pavement system, see also [44]. Feinberg [10] proved for various standard criteria, by using a convex-analysis approach, that for any policy there exists a nonrandomized Markov policy with the same or better performance. In the same paper, Feinberg [10], proved that property for mixtures of various criteria.

The 1990's saw systematic interest in criteria that mix several standard costs, which we now survey briefly. Krass, Filar and Sinha [29] studied a sum of a standard average cost and a standard discounted cost for models with finite state and action sets. They proved that for any $\varepsilon > 0$ there exists an ε-optimal randomized Markov policy which is stationary from some epoch N-onwards (a so-called ultimately deterministic or (N, ∞)-stationary policy). They also provided an algorithm to compute such policies. For the special case of a *communicating MDP*, the resulting policy is nonrandomized. As explained in Feinberg [11], the use of results from [10] simplifies the proofs in [29] and leads to the direct proof that for any $\varepsilon > 0$ there exists an ε-optimal (nonrandomized) Markov ultimately deterministic policy. The latter result can be derived from [29] but was formulated there only for the constant gain (communicating) case.

Fernandez-Gaucherand, Ghosh and Marcus [18] considered several weighted as well as overtaking cost criteria. They treat general (Borel) state and action spaces with bounded costs, and their objective is a linear combination of several discounted costs and an average cost. They prove that, for any $\varepsilon > 0$, there exists $N(\varepsilon)$ and an ε-optimal deterministic policy which coincides with the average-optimal policy after time $N(\varepsilon)$.

Ghosh and Marcus [20] consider a similar problem in the context of a continuous time diffusion model with positive costs. There is one average component and one discounted component. It is shown that it suffices to consider Markov (randomized) policies. Under global stability conditions they show the following. There is a minimizer in the class of stationary randomized controls and randomization is not necessary. For the more general minimization problem, there exist ε-optimal policies that agree with the optimal policy for the discounted part until a fixed time, and then switch to the policy which is optimal for the average part of the cost.

Hernández-Lerma and Romera [25] noticed that the "minimal pair" approach by Lasserre and Hernández-Lerma [23, 24] can be applied to multiple criteria problems when each criterion is of the form of either average rewards per unit time or total discounted rewards where different criteria may have different discount factors. The minimal pair approach (see Chapter 12) means that a controller selects a pair (μ, π) where μ is an initial distribution and π is a policy. In classical problems, the initial distribution is either fixed or a controller optimizes with respect to all initial states. Hernandez-Lerma and Romera [25] considered an additional condition which essentially means that π is randomized stationary and μ is an invariant distribution of the Markov chain with the transition probabilities defined by π. Naturally, for a one-step reward function r, in this case expected rewards at step t are equal to $\beta^{t-1}\mu r$, where β is the discount factor. Therefore, the average rewards per unit time are equal to μr and the expected total discounted rewards are equal to $\mu r/(1-\beta)$. Therefore, the change of the original formulation by using the minimal pair approach and the additional invariant condition on the initial measure μ reduce the problem with combined criteria (average rewards per unit time and discounted rewards) to a version of a problem with only average rewards per unit time.

Filar and Vrieze [19] considered a stochastic zero-sum game with finite state and action spaces. The objective is a linear combination: either of an average cost and a discounted cost, or of two discounted costs with two different discount factors. The immediate costs are the same for all components. They prove that in both cases, the value of the stochastic zero-sum game exists. Moreover, when the objective combines two discounted costs, both players possess optimal Markov policies. When the objective combines a discounted cost with an average cost, there exist ε-optimal (behavioral) policies.

Most of the papers on mixed criteria deal either with linear combinations of total discounted rewards with different discount factors or with a linear combination of total discounted rewards and average rewards per unit time. It appears that linear combinations of discounted rewards are more natural and easier to deal with than the weighted combinations of discounted and average rewards. The latter ones model the situation when there are two goals: a short-term goal modeled by total discounted rewards and a long-term goal modeled by average rewards per unit time. In the case of two different discount factors, the weighted discounted criterion models the same situation when one of the discount factors is close to 1. When the state and action sets are finite, optimal policies exist for mixed discounted criteria, they satisfy the Optimality Equation, and can be computed. Optimal policies may not exist for mixtures of total discounted rewards and average rewards per unit time [29]. Another advantage of dealing with mixed discounting is that for this criterion there is a well-developed theory for any finite number of discount factors [14, 15], while the papers that study linear combinations of discounted and average reward criteria usually deal with linear combinations of only two criteria: discounted rewards with a fixed discount factor and average rewards per unit time.

Single discount problems are often interpreted as total reward problems with a geometric life time: however, as shown in Shwartz [42], mixed discount problems cannot, in general, be interpreted as problems with several time scales.

In this paper we concentrate on a mixed-discounted problem for Markov decision processes. The exposition is based on the detailed study of this problem, performed by Feinberg and Shwartz [14, 15, 17].

In Section 7.2 we describe the model more precisely, and show through Example 7.1 that, although the weighted criterion seems like a small variation on a standard discounted problem, it induces quite different behavior on the resulting optimal policies. Then, in Section 7.3 we show that mixed-discounted problems can be reduced to standard discounted problems if we expand the state space $\mathbb{X}$ to $\mathbb{X} \times \mathbb{N}$, and that Markov policies are sufficient for one-criterion mixed-discounted problems. Section 7.4 obtains the characterization as well as an algorithm for the computation of optimal policies for the Weighted Discount Optimization problem (**WDO**) with finite state and action sets. **WDO** problems are introduced in Definition 7.1. In Section 7.5 we treat multiple-criterion problems and in Section 7.6 we discuss finite constrained problems. In Section 7.7 we survey existing results for other relevant problems, related models of stochastic games, and discuss extensions and open problems.