Approximate Dynamic Programming for the United States Air Force Officer Manpower Planning Problem

THESIS

MARCH 2017

Kimberly S. West, Captain, USAF

AFIT-ENS-MS-17-M-162

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-17-M-162

APPROXIMATE DYNAMIC PROGRAMMING FOR THE UNITED STATES AIR FORCE OFFICER MANPOWER PLANNING PROBLEM

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Kimberly S. West, BS, MS

Captain, USAF

MARCH 2017

AFIT-ENS-MS-17-M-162

APPROXIMATE DYNAMIC PROGRAMMING FOR THE UNITED STATES AIR FORCE OFFICER MANPOWER PLANNING PROBLEM

THESIS

Kimberly S. West, BS, MS
Captain, USAF

Committee Membership:

Lt Col Matthew J. Robbins, Ph.D.
Chair

Raymond R. Hill, Ph.D.
Member

AFIT-ENS-MS-17-M-162

Abstract

The United States Air Force (USAF) makes officer accession and promotion decisions annually. Optimal manpower planning of the commissioned officer corps is vital to ensuring a well-balanced manpower system. A manpower system that is neither over-manned nor under-manned is desirable as it is most cost effective. The Air Force Officer Manpower Planning Problem (AFO-MPP) is introduced, which models officer accessions, promotions, and the uncertainty in retention rates. The objective for the AFO-MPP is to identify the policy for accession and promotion decisions that minimizes expected total discounted cost of maintaining the required number of officers in the system over an infinite time horizon. The AFO-MPP is formulated as an infinite-horizon Markov decision problem, and a policy is found using approximate dynamic programming. A least-squares temporal differencing (LSTD) algorithm is employed to determine the best approximate policies. Six computational experiments are conducted with varying retention rates and officer manning starting conditions. The policies determined by the LSTD algorithm are compared to the benchmark policy, which is the policy currently practiced by the USAF. Results indicate that when the manpower system is in a starting state with on-target numbers of officers per rank, the ADP policy outperforms the benchmark policy. When the starting state is unbalanced, with more officers in junior ranking positions, the benchmark policy outperforms the ADP policy. When the starting state is unbalanced, with more officers in senior ranking positions, there is not statistical difference between the ADP and benchmark policy. In this starting state, ADP policy has smaller variance, indicating the ADP policy is more dependable than the benchmark policy.

To my boys, I love you infinity times infinity.

Acknowledgements

I would like to thank Dr. Robbins, my extremely patient and understanding advisor, for helping me through this academic journey. You, sir, truly pushed me beyond what I thought I could accomplish academically and displayed what it is to be an influential officer. Thank you.

I'd also like to thank Capt Phil Jenkins and MAJ Daniel Summers. You two helped me through this program, I can only hope I was able to return the favor in some small way.

Finally, a huge thank you to my family. You've been there through it all, and I love you for it.

Kimberly S. West

Table of Contents

List of Figures

List of Tables

APPROXIMATE DYNAMIC PROGRAMMING
FOR THE UNITED STATES AIR FORCE OFFICER
MANPOWER PLANNING PROBLEM

I. Introduction

1.1 Problem Background

The United States Air Force (USAF) provides national security capabilities in air, space, and cyberspace. A well-manned force is vital to carrying out its various missions. To maintain a competent and appropriately sized force, the USAF must attract and retain talented personnel [10]. This research seeks to improve policies regarding the management of the commissioned officer corps to support mission readiness.

Management of the commissioned officer corps is a manpower planning problem. In general, a manpower planning problem is determining the number of personnel, with specific skill sets, to best meet future operational requirements [18]. The USAF, along with its sister services, faces many challenges in manpower planning that a civilian organization does not. The most prominent challenge is the closed nature of the military. A closed manpower system is one in which new members can only join the organization at the lowest level. For the USAF, entrance for officers is only available at the rank of second lieutenant, or the O-1 grade. The only exceptions are in the medical, dental, and law fields. There are ten ranks (with corresponding grades) in the officer corps, starting at a second lieutenant (O-1) and ending at a general (O-10). A new general cannot be hired from outside the system; the individual can only be hired from the lieutenant general (O-9) pool. The hierarchical nature of the system is beneficial in ensuring uniformity of culture, which improves management [16], but creates difficulties in manpower planning.

Manpower planning must find a balance in meeting both short term and long term needs. Decisions made for the short term may have impacts that will not be realized for, potentially, 20 years to come. Currently, manpower planning is conducted by comparing historical attrition rates to current requirements for each Air Force Specialty Code (AFSC). An AFSC is an alphanumeric label for a specific career field within the USAF. Career groups (i.e., sets of career fields) include operations, logistics, support, medical, legal or chaplain, acquisition or finance, special investigation, special duty, and reporting [12]. Once the "optimal" number of officers is determined for each accession year group, a "sustainment line" is created for each AFSC. The sustainment line is used by manpower planning decision makers to determine how many Airmen are needed for each year group to sustain the career field over 30 years [35].

The USAF must not only aim to recruit the right number of talented people, but to keep those people. While the USAF is creating a more professional and technologically fluent force, the skill sets developed are making service-members more desirable to the private sector [16]. A balance must be found in recruitment and sustainment through promotion to avoid a system that does not have enough personnel or one that has too many. To manage the system, the USAF currently utilizes bonuses and reduction-in-force (RIF) mechanisms to maintain or reduce its size, respectively. These policy mechanisms cause career field managers to make constant and costly adjustments to the changes.

Manpower planning is a complex process wherein decisions must be made even though there is much uncertainty surrounding the decisions. In this study, a model is formulated to address the uncertainties in manpower planning. It will also provide insight concerning policies for USAF officer accession and sustainment. The research presented in this thesis addresses the following questions: (1) can the current manpower policy, with respect to accession and promotion, be improved? and (2) what is the impact of retention on manpower policy and the attendant costs? The Air Force Officer Manpower Planning Problem (AFO-MPP), developed by Bradshaw [8], is extended to study this important issue.

A Markov decision process (MDP) is constructed to model the AFO-MPP. An MDP models sequential decision making under uncertainty. It is comprised of decision epochs (or points in time), system states, available actions, state and action dependent immediate rewards or costs, and state and action dependent transition probabilities [33]. At each decision epoch, the decision maker (DM) chooses an action based on the system state. The result of the decision can provide the DM with a reward, and the system evolves to the next state. As this process continues, a sequence of rewards or costs is obtained. The objective is to maximize (minimize) expected total discounted reward (cost). A policy is attained that provides the DM a prescription for making decisions in the future [33].

In recent years, MDPs have become popular for solving manpower planning problems [27]. In a general Markov manpower planning model, a relationship between stock and flow (i.e., movement) of manpower is described over their variation in discrete time. Control over the system is typically accomplished through recruitment into the system or by varying rates of promotion [27].

In the AFO-MPP, the MDP has yearly decision epochs due to the yearly manpower authorization decisions made by Congress within the Department of Defense's (DoD) Future Years Defense Program [11]. The system states are defined by the number of officers in the system for each valid AFPC, rank, and commissioned years of service (CYOS) grouping. The two decision components include: (1) determining the number of officers to commission and bring into the system and (2) determining the number of officers to promote. Costs are based on the under-manned or over-manned status of each AFSC-rank combination. Finally, the transition probabilities reflect the uncertain number of officers remaining in the system after a decision is made. That is, they model the inherent stochasticity of retention rates (i.e., officers staying in the system). The objective is to minimize the expected total discounted under-manned and over-manned costs. The decision rule (i.e., manpower policy) indicates the number of officers to commission and the number of officers to promote for the appropriate AFSC-rank-CYOS combinations given the state of the manpower system.

Determining a stationary policy for realistically sized problems is computationally intractable. In order to address this "curse of dimensionality" [30], an approximate dynamic programming (ADP) algorithm is designed, developed, and tested to attain high-quality manpower planning policies relative to current practice. A design of experiment (DOE) is conducted to determine which ADP algorithm parameter settings produce the highest solution quality. Policies are sought that improve upon currently practiced USAF manpower planning policies.

1.2 Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 presents a detailed background of the sustainment problem, manpower planning models, and related operations research techniques. Chapter 3 describes the MDP model of the AFO-MPP and the ADP algorithm utilized to solve the model. Chapter 4 gives the computational experiments to evaluate quality of solutions attained by the ADP algorithm, as compared to currently practiced manpower policies. Chapter 5 concludes with a summary of the results and recommendations for future research.

II. Literature Review

The goal of manpower planning (MP) is to ensure "that the right people are available at the right places at the right time to execute corporate [Air Force] plans with the highest levels of quality," [24]. Various operations research (OR) techniques are applied to solve MP problems. In this chapter, prior research utilizing such techniques is reviewed. The focus, however, is on Markov decision processes (MDPs) and approximate dynamic programming (ADP). The development and application of MDPs to model many, smaller discrete stochastic problems is well documented [33, 43, 44, 45]. Unfortunately, as the size of the system increases, the ability to solve an MDP becomes computationally intractable. Thus, ADP techniques are applied to solve the MDP, as they can produce high-quality, implementable solutions [30]. Due to the relative newness of ADP, there are limited examples of previous work on solving MP problems. Instead, similar resource allocation problems are documented.

2.1 Operations Research to Solve Manpower Planning Problems

Wang [41] proposes the Training Force Sustainment Model (TFSM) for the Australian Army and reviews OR techniques typically applied to MP problems to inform development of his model. The OR techniques include MDPs, simulation, optimization, and system dynamics. The intention of the review is to reach beyond military MP. However, the specific characteristics of the military are discussed. For example, the military is closed in nature. That is, the military only recruits to fill the lowest rank and must fill higher ranks utilizing internal promotions. Also, the military utilizes both a *push-flow* and a *pull-flow* policy. A push-flow policy fills positions based on requirements such as an officer fulfilling a time requirement. A pull-flow policy fills positions through recruitment or promotion only when a position is available [41].

Simulation.

Simulation is a model of the real-world to imitate system behavior to approximate key characteristics or behaviors through the collection of statistics under given conditions [3]. The following papers serve as examples of MP research found in the literature that employ simulation. Onggo *et al*. [29] simulate the consequences of appraisal-system rules of promotion for the European Commission. The simulation model was created by the authors to demonstrate the likely consequences of various scenarios, not to produce predictions. This framework allowed comparison between the existing system and options for changing into a new system.

Blosch and Antony [7] combine computer simulation with experimental design to analyze the Royal Navy's manpower planning system. Simplified models are first built when developing simulations. The simplified model tests the major mechanism under study. Complex models are then gradually created to add the required level of accuracy on an agreed benchmark [7]. The strategic objectives for the Royal Navy include having accurate MP (i.e., ensure the Royal Navy has the right mix of specialties and grades to perform operational and support tasks required), effectively deploying manpower, managing careers, and giving advice to ministers.

Manpower planning is not limited to determining recruitment and promotion levels. Tang *et al*. [40] simulate manpower planning policies to facilitate cogent tactical and operational decisions concerning service territory size estimation and staffing level selection for after-sales field service (i.e., customer service). Mean response time, mean travel time, and max customer service representative utilization were estimated from the simulation runs to help evaluate the system design from different perspectives.

Optimization.

In general, optimization is concerned with finding the maxima (minima) of a real function from an allowable set of variables within a given problem [42]. Workman [46] develops a linear programming model to plan the generation of an indigenous security force over an unknown, infinite horizon. The Security Force Generation Model (SFGM) combines the growth of the enlisted and officer corps into one model, plans for growth over an infinite horizon, provides a variable-time planning horizon, and models the growth of the security force through recruitment, a legacy force, and enlisted accessions. Monthly and annual promotion rates are provided along with recruitment goals and accessions from the enlisted force. Data from the Afghan National Army is utilized.

Often, the objective of manpower planning models is to minimize the cost during recruitment and promotion periods. Cost is due to changes in the system. In Nirmala and Jeeva [28], the authors use dynamic programming through the Wagner-Whitin model to generate optimal recruitment and promotion schedules by minimizing the cost. Recruitment and promotion cohort sizes were assumed known and fixed; understaffing was not allowed; two grades were considered; and all costs were known.

Wu [47] proposes a fuzzy linear programming model for manpower allocation within a matrix organization. A matrix organization, or project, forms project teams within a line-staff organization. A project combines human and nonhuman resources to achieve a specific purpose and is assigned to a department. The management divisions seek to minimize costs under a limited manpower and project budget. The author's problem was modeled using fuzzy linear programming and solved with a two-phase approach. Phase one utilizes a max-min operator and phase two utilizes an average operator.

Companies seek to avoid high job turnover rates and want to prevent *brain-drain* of manpower [38]. To avoid this, predicting occupational life expectancy or mean residual life of those leaving is essential. In Sohn *et al*. [38], a random effects Weibull regression model is created to forecast these components. Both individual and non-individual characteristics are represented in the uncertainty. The authors test three hypotheses concerning turnover and potential reasons for turnover in Korean industry.

Organizations that do not operate on the typical 8-hour work schedule face demand fluctuations and find difficulty in optimizing the size and shape of their workforce. The United States Postal Service (USPS) main processing and distribution centers (P&DCs) face just this problem. Bard *et al*. [4] investigate the USPS P&DCs' demand fluctuations and workforce size and shape in two parts. First, they use historical data to analyze the demand distribution. Second, they develop and analyze a stochastic integer programming model to investigate potential end-of-month effects in demand. Full-time regular employees, part-time regular employees, and part-time flexible employees must meet the demand for a representative week during the year. The demand is deterministic and specified at fixed time increments over the baseline week. The authors were able to find a savings of about 4% by solving a two-stage recourse problem.

A multi-category workforce planning problem is addressed by Zhu and Sherali [48]. Functional areas located at different service centers, along with office space and recruitment capacity constraints, are modeled with fluctuating and uncertain workforce demand. A deterministic model is developed to accommodate fluctuations in expected demand. To address demand uncertainty, a two-stage stochastic process is proposed. The first stage makes recruiting and allocating decisions and the second stage reassigns workforce demand. This two-stage mixed-integer program is solved with a Benders' decomposition-based algorithm to minimize total costs.

The Army Medical Department (AMEDD) has a large number of medical specialties, making determining the number of hires and promotions for each specialty a complex task [5]. The authors introduce an objective force model (OFM), a deterministic, mixed-integer linear weighted goal-programming model to optimize manpower planning for AMEDD's medical specialist corp. Current practice is a manual approach that takes months to complete. The OFM uses discrete-event simulation to verify and validate the results of the deterministic model. The computational effort is reduced to seconds.

System Dynamics.

System dynamics (SD) takes a holistic approach in investigating complex dynamic behaviors of systems by analyzing structures and interactions of feedback loops, and time delays between actions and effects [41]. An *et al.* [2] create a workforce supply chain model by viewing project management as demand and viewing human resource management as supply. An SD modeling technique (i.e., systems thinking) is applied to find the causality relationships and feedback loops in the workforce supply chain. Two stocks (projects) are evaluated: proposed project and ongoing project. The input for the system is the inflow for the proposed project. The output is the execution rate. The decision rules are the number of people that should be hired during each period and how to handle skill evolution. Three feedback loops are captured in the model: a request rate, a corrected rate based on appeared skill gap, and an incorporated quantity of stock hired. The authors incorporate computer simulation to execute the SD model.

2.2 Markov Decision Processes

Markov chain theory is used to investigate dynamic behaviors of a system in a discrete time stochastic process wherein the evolution of the system, over time, is described by random variables [41]. Modeling MP to represent the stocks and flows of manpower lends naturally to the use of MDPs. A general representation of a discrete time Markov manpower system (MMS) is studied at discrete time epochs, consists of a finite number of grades, j, and the number of members in each grade is represented by a stochastic random variable, n_j. The transition probabilities represent the member moving to the next grade, through either promotion or reversion, or staying in the same grade [27]. Nilkantan and Roghavendra [27] make the distinction that while each individual may have their own unique probabilities of promotion, reversion, and remaining stationary, the behavior of the whole grade can be represented by the average behavior patterns of all individuals. The general MMS also models the number of recruits to each grade. The authors extend the general MMS model through control aspects in a hierarchical organization utilizing proportionality policies. The proportionality policies balance recruitment at every level but the entering level, with promotions at every level. The attainability, short term control, maintainability, and long term control were also discussed. The models incorporated with the proportionality restrictions are referred to as f-systems and f-models.

Nicholls [26] utilizes an MMS to analyze a graduate school in Australia's Doctor of Business Administration (DBA) program. The program was relatively new and needed information on expected success rates and expected first time passage to aid determination of supervisor workload. This system was viewed from both a short-term and long-term perspective. Candidates exit the system through an absorption state, either withdrawal or graduation. Data on 23 candidates was utilized to approximate the transition probabilities. The Markov chain was simplified due to the school's ability to have part time students and the closed nature of the system. In this situation, the closed nature of the system refers to the ability of a student to only enter the program in the first year.

Gans and Zhou [17] examine an employee staffing problem within a service organization. An MDP is created to capture the stochastic nature of employee learning and employee turnover. Three planning strategies are considered. The first strategy developed is for the long-term, high-level staffing problem. The second strategy looks at the medium-term, or mid-level workforce scheduling problem that uses Material Requirements Planning (MRP). The third strategy is for low-level work assignments that are viewed moment-by-moment. A telephone call center is described and modeled as a discrete-time, continuous-state space discounted MDP. The state variable represents the number of people at varying levels on the learning curve (i.e, gaining experience and speed through experience in handling calls). The MDP is solved via value iteration. The authors find that a *hire-up-to* policy is optimal under convexity, and a myopic policy is optimal otherwise.

The study by Dimitriou *et al*. [15] was motivated by the Greek debt reduction in 2012 and employed the Multivariate Non-Homogeneous Markov System (MNHMS). The MNHMS describes a manpower system through both horizontal and vertical mobility. The stocks (i.e., categories or departmental divisions) help determine the external recruiting and internal transfers of the system. Internal transfers are broken into intermobility, transitioning employees horizontally from one department to another, and intramobility, transitioning employees in the same department from one class to another. Fuzzy goal programming is utilized to model cost and reach the desired manpower structure.

Markov manpower systems tend to have the underlying assumption that they are time homogeneous and aperiodic. Gerontidis [20] provides a treatment to periodic Markov chains, specifically, periodicity in recruitment distribution and wastage probabilities. The parameters and recurrence equations describe the relative structures across time and the evolution of the expected grade.

Guerry [22] examines the problem of heterogeneity in a manpower system concerning both observable and unobservable variables. The author introduces a two-step procedure. The first step addresses homogeneous groups in terms of their transition probabilities and observable heterogeneity, such as age or gender. The second step considers heterogeneity in terms of unobservable sources and utilizes the mover-stayer principle. Movers are characterized by higher promotion probabilities and, therefore, have faster career growth. Stayers change their grade less frequently, if at all. A hidden Markov model is introduced to take into account the specifics of a manpower system and both the observable and latent sources of heterogeneity. A Markov-switching model is also used to model the phenomena of wastage and promotion flows.

In Blosch and Cantala [6], Markovian assignment rules are specified in terms of *agents* receiving *objects*. Both homogeneous and heterogeneous societies are analyzed. In the heterogeneous societies, agents are charactered by a pair (e.g., age and productivity). The four *natural* assignments considered are the seniority rule, the rank rule, the uniform rule, and the replacement rule. In the seniority rule, the older agents receive the object. In the rank rule, object j is assigned to the agent with object $j-1$. In the uniform rule, agents have equal probability in attaining the object. Finally, the replacement rule assigns an object to the entering agent. The transition probability matrices are computed over assignments generated by assignment rules.

Dimitriou and Tsantas [14] discuss the Generalized Augmented Mobility Model (GAMM). The GAMM is a Markov chain MP model that incorporates the use of training courses for existing employees to aid promotion, a preparation class for potential recruits outside of the organization, and the possibility that those same recruits could leave the preparation course before being hired. The manpower system is made up of an internal and external system. Internally, there are grades and training courses modeled by a non-homogeneous Markov chain. Externally, there is the preparation class.

The military must plan for manpower to meet current commitments but must also fulfill future political and military goals [18]. Gass [18] places individuals in descriptors, or classes. Under such large personnel systems, such as the military, it is difficult to track each person individually. Instead, it is convenient to place each person in a mutually exclusive class. The flows of personnel from one class to another are described through transition rates. These rates are used to forecast the next period of personnel inventories if given the correct initial class inventory.

The Army Manpower Long-Range Planning System (MLRPS) projects the United States Army strength for 20 years to develop long-range manpower plans using Markov chain-based approaches [19]. Gass *et al.* [19] break the problem into three subsystems: the data processing subsystem, the flow model subsystem, and the optimization subsystem. In the data processing subsystem, data is collected to generate historical and projected rates. The rates become the input for the flow subsystem that uses a Markov chain model to project the flow of the initial force over a 20-year time horizon. The output of the Markov chain is the input to the optimization subsystem.

Škulj *et al.* [37] utilize Markov chain models to aid in attainability and maintainability of manpower in the Slovenian armed forces. The authors' transition probabilities modeled recruitment into Slovenian military segments. They divided the Slovenian population into 126 segments: six general or non-military and 120 military based on their administrative title. The authors quickly identify the difficulty in solving the MDP model due to the large size and numerous possible transitions. The authors found all the possible transitions to be challenging to implement in such a large model. The next section discusses the use of approximate dynamic programming as a method for solving large-scale MDPs.

2.3 Approximate Dynamic Programming

When a dynamic program is computationally intractable, approximate dynamic programming (ADP) can be used to solve the so called "curses of dimensionality" [30]. Dimensionality difficulties can occur in the state space, outcome space, and/or action space. Any combination of these only makes the problem more difficult. Thus, the techniques applied in ADP can attain approximate solutions to problems that have state, outcome, or action variables with millions of dimensions. ADP algorithms have been shown to produce quality solutions, at times within one percent of optimal. The algorithms utilize Monte Carlo simulation to sample random outcomes in both the state and action spaces in order to determine the value of the outcome.

When using approximation techniques, a balance between computational efficiency and the performance of the resulting policy must be taken into consideration [31]. Further, the architecture of the approximation may aid in overcoming computational challenges. A specially structured approximation architecture could ease the challenges. For example, linear and separable concave architectures have been successfully applied to a variety of problems in transportation operations [31].

Song and Huang [39] use the successive convex approximation method (SCAM) to solve a multistage stochastic MP problem. They utilize a piecewise linear and convex function to approximate the value function. The authors seek to plan for the transferring, hiring, or firing of employees among different branches of an organization with uncertain workforce demand and turnover. The authors were able to solve the MP problem within 0.02% of the optimal solution, on average.

ADP algorithms and techniques have been applied to Air Force Officer Manpower Planning Problem (AFO-MPP) through the works of Hoecherl *et al*. [23] and Bradshaw [8]. Hoecherl *et al*. develop two ADP algorithms to consider accessions and promotion decisions for multiple AFSCs, officer grades, and year groups. First, the authors apply least-squares approximate policy iteration (LSAPI) to determine approximate policies. The algorithm employs a modified version of the Bellman equation based on the post-decision state variable. Second, an approximate value iteration algorithm, a variant of the concave adaptive value estimation algorithm, is developed to identify an improved policy for the current USAF officer sustainment system [21]. In Bradshaw [8], a single AFSC and accessions-only decisions are considered. Here, an LSAPI is also used to attain solutions. Two MP problem instances are created to compare to the performance of the ADP technique to a benchmark policy.

Due to the relative newness of ADP, there are limited examples of previous work on solving MP problems. The following examples are similar resource allocation problems. Ahner and Parson [1] consider the optimal allocation of weapons to a collection of targets. The objective is to maximize the reward for destroying the targets. The problem has two stages. In the first stage, targets are known. In the second stage, targets arrive in a random distribution. The authors utilize a solution approach for the dynamic weapon target assignment (DWTA) problem that involves Monte Carlo sampling to solve the DWTA. The authors were able to solve to optimality using the approximation.

Rettke *et al*. [34] formulate an MDP to examine a military medical evacuation (MEDEVAC) dispatching problem. The ADP approximate policy iteration algorithmic strategy of least squares temporal differences (LSTD) is utilized. A representative planning scenario is created to compare the ADP policy to the myopic policy. The ADP policy performs up to 31% better than the myopic. Similarly, Davis *et al*. [13] sought to optimize a defensive response to a missile attack using LSTD. The four instances tested indicate that the ADP policy uses minimal computation effort and achieves a 7.74% optimality gap over the computationally heavy MDP optimal solution.

Schneider National, the largest truckload motor carrier in the United States, in collaboration with CASTLE Laboratory at Princeton University, develop a model to answer a myriad of questions concerning hiring, estimating work rule changes, managing drivers, and experimenting with new routes [36]. The model seeks to optimize decisions over time regarding driver allocation to varying loads with different load characteristics. A state of the resource is defined by an attribute vector that is comprised of attributes such as location, domicile, capacity type (i.e., a team of drivers, a solo driver, or an independent contractor), and others. The loads are also assigned attributes that aided in determining costs. Each decision indicates whether the truck should be moved with a load or moved empty with anticipation that the new location will yield a greater contribution. The problem is solved by breaking it into two time stages. A pre-decision state is computed depending on the post-decision state. Further, the authors perform an ADP double-pass algorithm in which they simulate decisions forward in time without updating the value functions; the derivative is then computed in a backward pass. Schneider required the authors' model to match historical data within a specified range. The results closely matched historical data [36].

Using data from Canadian military airlift operations, Powell *et al*. [32] examine the impact of uncertain customer demands and aircraft failure on cost. Myopic policy decisions are first analyzed and then compared to results obtained through ADP. A myopic policy is the most elementary policy that does not use forecasted information or any direct representation of decisions in the future [30]. The authors utilize ADP to produce robust decisions that are less sensitive to uncertainty. They are able show that their robust decisions perform better than the myopic policy by reducing the value of *advance information*.

III. Methodology

3.1 Problem Statement

The Air Force Officer Manpower Planning Problem (AFO-MPP) from Bradshaw [8] is extended to include officer promotion decisions. The decisions concerning how many officers to hire and how many officers to promote must be made sequentially over time and under uncertainty. The system level uncertainty results from individual officer retention outcomes - officers may elect to remain in the system or to exit the system by separating or retiring. Since current manpower decisions affect the state and cost of the personnel system in the future, the impact of current hiring and promotion decisions on the future state of the system must be considered. As such, a Markov decision process (MDP) model of the AFO-MPP is formulated. The objective of the MDP is to identify the manpower policy (i.e., hiring and promotion decisions as a function of the state of the personnel system) that minimizes the expected total discounted cost of maintaining the required number of officers in the system over an infinite horizon.

For the formulation of the AFO-MPP as an MDP, the state space indicates the number of officers in the system within a specific Air Force specialty code (AFSC), rank, and commissioned years of service (CYOS) combination. The tuple of AFSC-rank-CYOS is chosen for the state space to reflect the career field, the military pay grade, and the number of years officers have been in the system. The state space is limited by only considering officers of the grades O-1 through O-6. General officers, grades O-7 through O-10, are not considered due to their unique promotion system. The action space captures how many officers to bring into the system, or how many O-1 officers to access, and how many officers to promote from one rank to the next. Officers accessed or promoted at time period t are assumed ready for duty at that time. The accession process is simplified because officers can be commissioned at different points throughout the year due to varying commissioning sources. Similarly, the promotion process is simplified because officers can be promoted during promotion boards that take place throughout the year. Officers that are neither assessed nor promoted will be transitioned from one time period to the next, taking into account the random retention rates of officers throughout the time period. The retention rates of officers with different AFSCO-rank-CYOS attributes may differ. The event timing diagram in Figure 1 displays the transition of officers through the AFO-MPP MDP model, utilizing notation that will be introduced in the subsequent section.

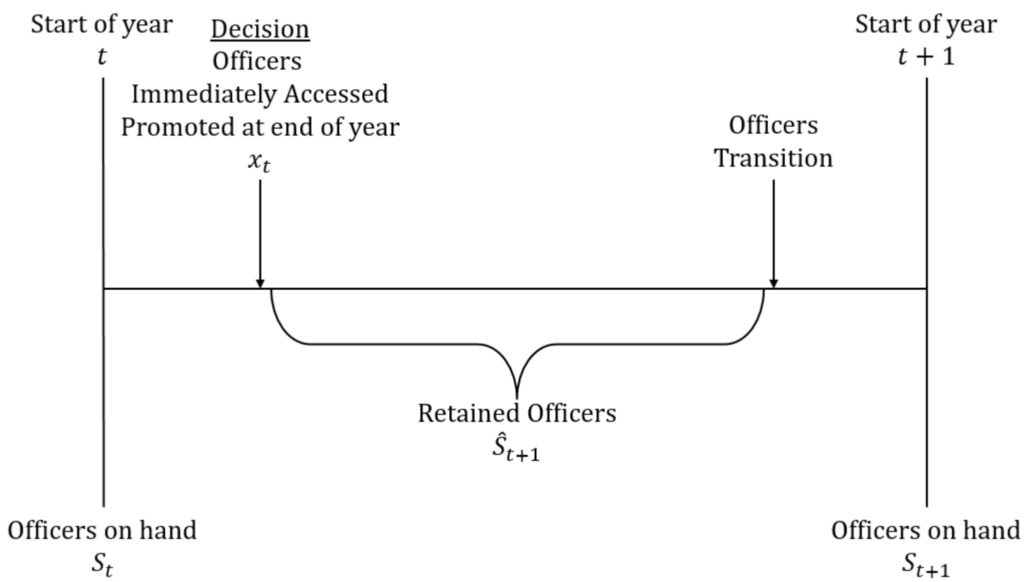


Figure 1. Event Timing Diagram for AFO-MPP

The costs for the AFO-MPP are due to over-manning and under-manning within the system. The military must plan for manpower to meet current requirements, but must also fulfill future requirements [18]. When the requirements are not met, a cost is incurred. The objective of the MDP is to select a manpower policy that minimizes expected total discounted cost. A manpower policy is a decision rule that indicates how many officers to access into the USAF and how many officers to promote given the current state of the system. The AFO-MPP is formulated as a discrete-time, discrete-state MDP.

3.2 MDP Formulation

The MDP model for the AFO-MPP is described as follows:

Decision Epochs

Decisions are made annually about the hiring of new officers into the system and promoting of officers to the next rank. The set of decision epochs is denoted as follows:

$$\mathcal{T} = \{0, 1, 2, \ldots\}. \tag{1}$$

The epochs are the points in time at which decisions are made. In this case, the accession and promotion decisions are made at the beginning of the year.

State Space

The state space of the system is comprised of the number of officers with selected AFSC, rank, and CYOS attribute combinations. For this thesis, we propose an aggregate officer replacement model wherein we express the AFSC, rank, and CYOS of the officer. Let

$$a \in \mathcal{S}^{AFSC} = \text{ AFSC of an officer,}$$

where $\mathcal{S}^{AFSC} = \{1, 2, \ldots, A\}, A < \infty$, denotes the set of AFSCs. Let

$$r \in \mathcal{S}^{rank} = \text{ rank of an officer,}$$

where $\mathcal{S}^{rank} = \{1, 2, \ldots, 6\}$ denotes the set of ranks, representing officer ranks O-1 through O-6. Let

$$y \in \mathcal{S}^{CYOS} = \text{ CYOS of an officer,}$$

where $\mathcal{S}^{CYOS} = \{1, 2, \ldots, 29\}$ denotes the set of CYOS. It is possible for an officer to enter the system with enlisted years of service. For the purpose of this thesis, only commissioned service is considered.

The set $\mathcal{S}$ contains the full scope of possible combinations of AFSC a, rank r and CYOS y. Let

$$(a, r, y) \in \mathcal{S} = \text{set of all possible officer AFSC-rank-CYOS attribute combinations,}$$

where $\mathcal{S} = \mathcal{S}^{AFSC} \times \mathcal{S}^{rank} \times \mathcal{S}^{CYOS}$. Not all combinations are feasible due to the hierarchical nature of the system. For example, an O-1 would not have 25 CYOS. Figure 2 indicates the feasible combinations.

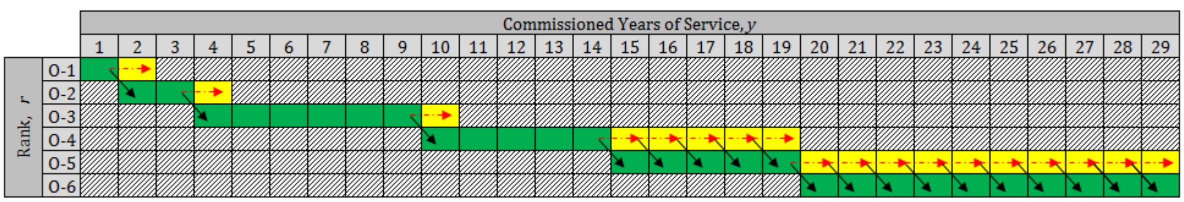


Figure 2. Feasible rank-CYOS combinations

The state of the system is determined by the number of officers in AFSC a, of rank r, and with CYOS y, $\forall\ (a, r, y) \in \mathcal{S}$. Let

$$S_{tary} = \text{the number of officers at time } t \text{ of AFSC } a \text{, rank } r \text{, and CYOS } y. \tag{2}$$

Note that $S_{tary} \in \mathbb{N}^0$, $\forall\ (a, r, y) \in \mathcal{S}$ cannot take on negative values because of the nature of the personnel system. The pre-decision state is a vector denoted as

$$S_t = (S_{tary})_{(a,r,y) \in \mathcal{S}}.$$

Action Sets

The action at time t indicates the number of officers to be commissioned into AFSC a at rank $r = 1$ and the number of officers to be promoted from AFSC a, rank r, and CYOS y. Let

$$x_t = (x_t^{access}, x_t^{promote}) \tag{3}$$

where

$$x_t^{access} = (x_{ta}^{access})_{a \in \mathcal{S}^{AFSC}}$$

and

$$x_{ta}^{access} \in \mathbb{N}^0,$$

is the number of officers commissioned into AFSC a at rank $r = 1$ at time t. Officers entering the USAF at rank O-1 (i.e., no prior experience) are considered to have zero CYOS. Let

$$x_t^{promote} = (x_{tary}^{promote})_{(a,r,y) \in \mathcal{S}^{promote}},$$

where

$$\mathcal{S}^{promote} = \cup_{a \in \mathcal{S}^{AFSC}} \Big\{ (a,1,1), (a,2,3), (a,3,9), (a,4,14), (a,5,19) \Big\} \subseteq \mathcal{S}$$

and

$$x_{tary}^{promote} \in \{0, 1, \ldots, S_{tary}\}$$

is the number of officers promoted from AFSC a, rank r, and CYOS y to AFSC a, rank $r+1$, and CYOS $y+1$, effective at time $t+1$.

Transition Probabilities

Attributes AFSC a, rank r, and CYOS y all influence the probability of retention of a USAF officer. Let

$$\psi_{ary} = \text{the probability an officer with attribute tuple } (a, r, y) \in \mathcal{S}$$
$$\text{will be retained in the system (i.e., does not separate or retire).}$$

The probability of retention may differ for each AFSC-rank-CYOS combination. $\hat{S}_{t+1,ary}$ is a random variable following a binomial distribution with parameters S_{tary} and ψ_{ary}. Stated simply, the number of officers with attribute combination (a, r, y) remaining in the system at time $t+1$ depends on the number in the system at time t.

The number of officers of AFSC a, rank r, and CYOS y available at time $t+1$, $S_{t+1,ary}$, results from the number of officers of AFSC a, rank r, and CYOS $y-1$ in the system at time t, $S_{t,a,r,y-1}$; the number of new officers accessed at time t, x_t^{access}; the number of officers of AFSC a, rank r, and CYOS $y-1$ that retain during the time interval $(t, t+1)$, $\hat{S}_{t+1,a,r,y-1}$; and officer promotions, $x_t^{promote}$. Officer promotions occur only during specific promotion windows (as indicated by $S^{promote}$). Refer to Figure 2 to visualize promotion windows.

Retention, accessions, and promotions are modeled via the following transition function:

$$S_{t+1,ary} = \begin{cases} \hat{S}_{t+1,ary}(x^{promote}_{t,a,r-1,y-1}; \psi_{a,r-1,y-1}) & \text{if } (a, r-1, y-1) \in \mathcal{S}^{promote} \\ \hat{S}_{t+1,ary}(S_{tar,y-1}; \psi_{ta,r,y-1}) & \text{if } (a, r, y-1) \in \mathcal{S} \setminus \mathcal{S}^{promote} \\ x^{access}_{ta} & \text{if } (a, r, y) = (a, 1, 1) \\ \hat{S}_{t+1,ary}(S_{ta,r,y-1} - x^{promote}_{t,a,r,y-1}; \psi_{a,r,y-1}) & \text{if otherwise.} \end{cases} \tag{4}$$

The first case denotes officers within a promotion window who will either promote and retain in the system or leave the system. The second case denotes officers not in a promotion window. The third case denotes newly accessed officers. The fourth case denotes officers who are within the promotion window but do not promote. The transition of officers in the AFO-MPP can be written in the following system dynamics form:

$$S_{t+1} = S^M(S_t, x_t, \hat{S}_{t+1}).$$

Costs

The cost for the AFO-MPP is due to over-manning and under-manning within the personnel system. The military must plan for manpower to meet current requirements, but must also fulfill future requirements as well [18]. When the requirements are not met, a cost is incurred. To model the requirements, a value $\bar{S}_{ar}$ is specified that indicates the number of officers required for each AFSC a and rank r. Let $c^o_{ar} > 0$ denote the cost of an over-manned rank and $c^u_{ar} > 0$ denote the cost of an under-manned rank for each AFSC a and rank r. The over-manned cost function for each AFSC a and rank r combination is

$$O_{ar}(S_t, x_t) = \begin{cases} c^o_{ar}\Big(max\{(x_{ta} + \sum_{y\in\mathcal{S}^{CYOS}} S_{tary}) - \bar{S}_{ar}, 0\}\Big) & \text{if } r = 1, \\ c^o_{ar}\Big(max\{(\sum_{y\in\mathcal{S}^{CYOS}} S_{tary}) - \bar{S}_{ar}, 0\}\Big) & \text{if } r > 1. \end{cases} \tag{5}$$

There is a separate cost function for O-1 officer surplus because officers accessed during time period t must be included. Officers of ranks O-2 through O-6 are modeled the same. The under-manned cost function for each AFSC a and rank r combination is

$$U_{ar}(S_t, x_t) = \begin{cases} c^u_{ar}\Big(max\{(\bar{S}_{ar} - (x_{ta} + \sum_{y\in\mathcal{S}^{CYOS}} S_{tary}), 0\}\Big) & \text{if } r = 1, \\ c^u_{ar}\Big(max\{\bar{S}_{ar} - (\sum_{y\in\mathcal{S}^{CYOS}} S_{tary}), 0\}\Big) & \text{if } r > 1. \end{cases} \tag{6}$$

Again, there is a separate cost function for O-1 officer shortages because officers accessed during time period t must be included. Officers of ranks O-2 through O-6 are modeled the same. Thus, a single period cost function for the AFO-MPP is the sum of the over-manned and under-manned costs. Utilizing Equations 5 and 6, the cost function is

$$C(S_t, x_t) = \sum_{(a,r)\in\mathcal{S}^{AFSC}\times\mathcal{S}^{rank}} O_{ar}(S_t, x_t) + U_{ar}(S_t, x_t). \tag{7}$$

Objective Function

Having described all components of the MDP model, through Equations 1, 2, 3, 4, and 7, the objective for the AFO-MPP can be formulated as:

$$\min_{\pi\in\Pi} \mathbb{E}\Bigg\{\sum_{t=0}^{\infty} \gamma^t C(S_t, x_t)\Bigg\}. \tag{8}$$

Identifying a policy $\pi \in \Pi$ that minimizes the expected total discounted cost is the goal of this thesis. The cost is discounted through use of γ. The optimal policy provides the number of second lieutenants to hire and the number of officers of grade O-2 through O-6 to promote in each rank during each time period that would be the least costly. Solving Bellman's equation provides the optimal policy:

$$V(S_t) = \min_{x_t}\left(C(S_t, x_t) + \gamma\mathbb{E}\Big\{V(S^M(S_t, x_t, \hat{S}_{t+1}))|S_0\Big\}\right).$$

3.3 Approximate Dynamic Programming Algorithms

Approximate Policy Iteration

Due to its high dimensionality, the AFO-MPP is solved using an ADP technique. ADP provides a mechanism to approximate the value function without having to enumerate the state space and compute the value of each state-action pair. Monte Carlo simulations are employed in ADP algorithms to sample the random outcomes in the state and action spaces and determine the value of these outcomes. If a sufficiently large state space is sampled, some ADP algorithms are shown to converge to optimality [30]. The use of Monte Carlo simulation alleviates the need to solve for the value of each state-action pair. Utilization of the post-decision state convention can reduce the computational complexity in ADP algorithms [30]. The post-decision state, S_t^x, considers the state of the system immediately prior to the revelation of exogenous changes to the system, allowing the expectation to be computed outside of the minimization operator. Bellman's equation is represented as follows when the post-decision state is incorporated:

$$V^x(S_{t-1}^x) = \mathbb{E}\Big\{\min_x C(S_t, x) + \gamma V^x(S_t^x)|S_{t-1}^x\Big\}. \tag{9}$$

Approximate policy iteration (API) is an ADP algorithmic strategy that evaluates the values associated with states and outcomes for a fixed policy, or set of actions, for an MDP problem. API's benefit is the ease with which values of policies are found [30]. The policy iteratively updates based on the the observed values of the fixed policy. API with parametric modeling and linear basis functions allow linear regression techniques to be applied to estimate a parameter vector θ to fit a value function approximation using the selected basis functions. The value function approximation utilized in the AFO-MPP API implementation leverages the post-decision state variable convention and is denoted below:

$$\bar{V}^x(S_t^x) = \sum_{f \in \mathcal{F}} \theta_f \phi_f(S_t^x) = \theta^\top \phi(S_t^x), \tag{10}$$

where $\Big(\phi_f(S_t^x)\Big)_{f \in \mathcal{F}}$ is the set of basis functions for the post-decision state. Substituting this equation into Bellman's equation gives the foundation for least squares value function approximation:

$$\theta^T \phi(S_{t-1}^x) = \mathbb{E}\Big\{C(S_t, X^\pi(S_t|\theta)) + \gamma\theta^\top \phi(S_t^x)|S_{t-1}^x\Big\} \tag{11}$$

where X^π is the policy function for the MDP model.

Inner Minimization Problem

Even with the dimensionality reduction attained by using the post-decision state variable, the optimality equation remains computationally intractable due to the high dimensionality of the feasible action space. To determine the policy function, as given by X^π, an inner minimization problem (IMP) is formulated and solved to determine which action, x_t, should be taken. The IMP is first defined as a non-linear integer program.

$$
\begin{aligned}
X^{\pi}(S|\theta) = \underset{x}{\operatorname{argmin}} & \sum_{a\in\mathcal{S}^{AFSC}} \left[\left| x^{access} + S_{a,1,1} + S_{a,1,2} - \bar{S}_{a,1} \right| \right] + \gamma \sum_{f\in\mathcal{F}} \theta_f \phi_f \\
s.t. \quad & x^{promote}_{a,1,1} \leq S_{a,1,1}, \forall a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,2,3} \leq S_{a,2,3}, \forall a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,3,9} \leq S_{a,3,9}, \forall a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,4,15} \leq S_{a,4,15}, \forall a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,5,19} \leq S_{a,5,19}, \forall a \in \mathcal{S}^{AFSC} \\
& x^{access}_{a} \in \mathbb{N}^0, \forall\ a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,r,y} \in \mathbb{N}^0,\ \forall\ (a,r,y) \in \mathcal{S}^{promote}
\end{aligned}
$$

The t subscript is dropped for notational simplicity. To obtain a tractable approach, the following five linear basis functions are developed. Moreover, only one AFSC is considered to simplify exposition of the approach. The basis functions ϕ_f, $f = 1, 2, \ldots, 5$, are defined as:

$$
\begin{aligned}
\phi_1 &= \left| \psi_{a,1,1} x^{promote}_{a,1,1} + \psi_{a,2,2} S_{a,2,2} + \psi_{a,2,3}(S_{a,2,3} - x^{promote}_{a,2,3}) - \bar{S}_{a,2} \right|, \\
\phi_2 &= \left| \psi_{a,2,3} x^{promote}_{a,2,2} + \sum_{y=4}^{8} \psi_{a,3,y} S_{a,3,y} + \psi_{a,3,9}(S_{a,3,9} - x^{promote}_{a,3,9}) - \bar{S}_{a,3} \right|, \\
\phi_3 &= \left| \psi_{a,3,9} x^{promote}_{a,3,9} + \sum_{y=10}^{13} \psi_{a,4,y} S_{a,4,y} + \psi_{a,4,14}(S_{a,4,14} - x^{promote}_{a,4,14}) - \bar{S}_{a,4} \right|, \\
\phi_4 &= \left| \psi_{a,4,14} x^{promote}_{a,4,14} + \sum_{y=15}^{18} \psi_{a,5,y} S_{a,5,y} + \psi_{a,5,19}(S_{a,5,19} - x^{promote}_{a,5,19}) - \bar{S}_{a,5} \right|, \\
\phi_5 &= \left| \psi_{a,5,19} x^{promote}_{a,5,19} + \sum_{y=20}^{28} \psi_{a,6,y} S_{a,6,y} - \bar{S}_{a,6} \right|.
\end{aligned}
$$

The post-decision state is implicit in the formulation of the basis functions (i.e., $S^{M,x}(S_t, x_t) = S_{a,r,y} - x^{promote}_{a,r,y}\ \forall (a, r, y) \in \mathcal{S}^{promote} \backslash (a, 1, 1))$. The basis functions define the number of officers under and over the target value of officers, $\bar{S}_{a,r}$, for each AFSC $a \in \mathcal{S}^{AFSC}$ and rank r, $r = 2, 3, \ldots, 6$. For example, ϕ_1 is the number of second lieutenants promoting to first lieutenant, the number of first lieutenants remaining first lieutenant, and removes the number of first lieutenants promoting to captain. The absolute value of the officers remaining in the first lieutenant rank, minus the target value, determines the number of officers under or over.

The IMP is then transformed from a non-linear integer program into a linear integer program by replacing each absolute value with a new decision variable, z.

$$
\begin{aligned}
X^{\pi}(S|\theta) = \underset{z,x}{\operatorname{argmin}} \quad & z_1 + \gamma\Big(\theta_1 z_2 + \theta_2 z_3 + \theta_3 z_4 + \theta_4 z_5 + \theta_5 z_6\Big) \\
s.t. \qquad & z_1 \geq \pm x^{access} + S_{a,1,1} + S_{a,1,2} - \bar{S}_{a,1} \\
& z_2 \geq \pm\gamma\theta_1(\psi_{a,1,1} x^{promote}_{a,1,1} + \psi_{a,2,2} S_{a,2,2} + \psi_{a,2,3}(S_{a,2,3} - x^{promote}_{a,2,3}) - \bar{S}_{a,2}) \\
& z_3 \geq \pm\gamma\theta_2(\psi_{a,2,3} x^{promote}_{a,2,2} + \sum_{y=4}^{8} \psi_{a,3,y} S_{a,3,y} + \psi_{a,3,9}(S_{a,3,9} - x^{promote}_{a,3,9}) - \bar{S}_{a,3}) \\
& z_4 \geq \pm\gamma\theta_3(\psi_{a,3,9} x^{promote}_{a,3,9} + \sum_{y=10}^{13} \psi_{a,4,y} S_{a,4,y} + \psi_{a,4,14}(S_{a,4,14} - x^{promote}_{a,4,14}) - \bar{S}_{a,4}) \\
& z_5 \geq \pm\gamma\theta_4(\psi_{a,4,14} x^{promote}_{a,4,14} + \sum_{y=15}^{18} \psi_{a,5,y} S_{a,5,y} + \psi_{a,5,19}(S_{a,5,19} - x^{promote}_{a,5,19}) - \bar{S}_{a,5}) \\
& z_6 \geq \pm\gamma\theta_5(\psi_{a,5,19} x^{promote}_{a,5,19} + \sum_{y=20}^{28} \psi_{a,6,y} S_{a,6,y} - \bar{S}_{a,6}) \\
& x^{promote}_{a,1,1} \leq S_{a,1,1} \\
& x^{promote}_{a,2,3} \leq S_{a,2,3} \\
& x^{promote}_{a,3,9} \leq S_{a,3,9} \\
& x^{promote}_{a,4,15} \leq S_{a,4,15} \\
& x^{promote}_{a,5,19} \leq S_{a,5,19} \\
& x^{access}_{a} \in \mathbb{N}^0, \forall\, a \in \mathcal{S}^{AFSC} \\
& x^{promote}_{a,r,y} \in \mathbb{N}^0,\ \forall\, (a,r,y) \in \mathcal{S}^{promote} \\
& z_i \geq 0, i = 1, 2, \ldots, 6
\end{aligned} \tag{12}
$$

Least Squares Temporal Differences

The least squares temporal differencing algorithm is an on-policy algorithm that minimizes the sum of the temporal differences, or Bellman's error, for approximating the estimation of the true value function [30]. Minimizing Bellman's error minimizes the difference between the approximation of the value function and the observed value of the approximations. The estimator for the least squares Bellman error minimization is as follows:

$$\hat{\theta} = [(\Phi_{t-1} - \gamma\Phi_t)^\top(\Phi_{t-1} - \gamma\Phi_t)]^{-1}(\Phi_{t-1} - \gamma\Phi_t)^\top C_t, \tag{13}$$

where Φ_{t-1} is a matrix of basis function evaluations for the sampled post-decision states, Φ_t is a matrix of basis function evaluations for the sampled post-decision states in the next period, and C_t is a vector of observed costs of the period in each iteration. θ is smoothed, allowing algorithm convergence to slow [30], utilizing generalized harmonic smoothing. The formula for smoothing is is as follows:

$$\frac{a}{a+n-1}. \tag{14}$$

Bradtke and Barto [9] first introduced LSTD. A variant, utilizing value function approximations around post-decision states, as recommended by Powell [30], is outlined in Algorithm 1.

Algorithm 1 API Algorithm

Step 0: Initialize θ^0.
Step 1:
for n=1 **to** N **(Policy Improvement Loop)**
  Step 2:
  for m=1 **to** M **(Policy Evaluation Loop)**
    Generate a random post-decision state, $S^x_{t-1,m}$.
    Record basis function evaluation $\phi(S^x_{t-1,m})$.
    Simulate transition to next epoch, obtain a pre-decision state, $S_{t,m}$.
    Determine decision $x = X^\pi(S_{t,m}|\theta^{n-1})$ by solving inner minimization problem (IMP).
    Compute post-decision state $S^x_{t,m}$.
    Record cost $C(S_{t,m}, x_t)$.
    Record basis function evaluation $\phi(S^x_{t,m})$
  end for
  End
  Update θ^n using Equation 13 and Equation 14.
end for
Return $X^\pi(S_t|\theta^N)$ and θ^N.
End

A Latin hypercube sampling (LHS) technique is used to generate the random sample of post-decision states for Step 2 of Algorithm 1. A benefit in using LHS is that it allows for uniform sampling across all dimensions.

IV. Computational Results

This chapter applies the approximate dynamic programming (ADP) techniques outlined in Chapter 3 to the U.S. Air Force officer manpower planning problem (AFO-MPP). ADP algorithm features are investigated to determine the impact on solution quality. An experimental design is conducted in MATLAB to determine which features produce the most superior results as compared to the benchmark policy. Six scenarios for the AFO-MPP are investigated. An experimental design is executed for each scenario to test the performance of the ADP algorithm against the benchmark policy and determine the computational effort required.

Benchmark Policy.

The United States Air Force (USAF) currently determines accession rates utilizing retention rates in comparison to the desired force end strength. The goal is to access a number of officers every year, over a thirty year time horizon, to maintain the desired end strength. If the current state of the system is under-manned, the number of officers needed to maintain end strength, plus the gap in force, are accessed. For example, consider a situation wherein the desired end strength is 650 officers and the retention rates indicate that 28 officers should be accessed every year for thirty years to maintain the force. If the current state of the system has 610 officers, 40 additional officers should be accessed, for a total of 68 officers. If the current state of the system is over-manned, just the number of officers needed to maintain end strength are accessed. For example, if the current state of the system has 660 officers, only the 28 officers needed to maintain a 650 end strength force are accessed.

Experimental Design.

A set of experiments is constructed to evaluate the proposed ADP algorithm's solution quality and computational effort by studying the impact of systematically varying different features of the ADP algorithm and certain Markov decision process (MDP) parameters [25]. The policy resulting from the ADP algorithm is assessed based on its improvement over a benchmark policy for the AFO-MPP. The benchmark policy is defined based on information provided by Headquarters Air Force (HAF-A1) on the 61A Career Field. The response variable for the design of experiments is the mean total discounted cost of the ADP policy and the mean total discounted cost of the benchmark policy. The half-width for the mean cost is reported at the 95% confidence level. The computation times for the ADP algorithm are recorded to measure the computational effort needed to perform the ADP algorithm.

Four algorithmic features for the AFO-MPP are investigated. A fifth feature, the utilization of instrumental variables, or solely employing the Bellman error minimization, was screened out from the final design. The presence of instrumental variables increases the cost of the ADP policy in all situations as compared to the cost without the use of instrumental variables. The first algorithmic feature considered is the number of policy improvement (outer) loops (N), set to 25 and 50. The second feature considered is the number of policy evaluation (inner) loops (M), set to 1,000 and 5,000. The third feature considered is a regularization parameter (η), set at 10 and 100. The final feature examined is the parameter for the generalized harmonic smoothing function, a, found in Equation 14. The low factor setting for a is $a = 1$ to study how simple harmonic smoothing affects the response. The high factor setting for a is $a = 10$. This allows the algorithm convergence to slow [30]. Table 1 summarizes the algorithmic features and their levels.

A 2^4 full factorial design with five replicates is implemented. One replicate can

Table 1. Design Factor Settings

Factor	Low	High
Policy Improvement (N)	25	50
Policy Evaluation (M)	1000	5000
Regularization (η)	10	100
Harmonic Smoothing (a)	1	10

be seen in Table 2. All terms are free from aliasing. This design investigates the four selected features in five replicates, for a total of 80 runs. When applying this experimental design, the ADP policy is created by calculating the θ coefficients for the basis functions from the implementation of the ADP algorithm. Each of the 80 runs resulted in a different θ coefficient vector. Once obtained, the θ coefficients are utilized to conduct a simulation for both the ADP policy and the benchmark policy over a 30 year horizon for 30 replications per treatment to obtain the statistics of the response variables. Common random numbers (CRN) were utilized to reduce variance.

Table 2. Full Factorial Replicate

N	M	η	a
25	1000	10	1
25	1000	10	10
25	1000	100	1
25	1000	100	10
25	5000	10	1
25	5000	10	10
25	5000	100	1
25	5000	100	10
50	1000	10	1
50	1000	10	10
50	1000	100	1
50	1000	100	10
50	5000	10	1
50	5000	10	10
50	5000	100	1
50	5000	100	10

Experimental Results.

The different scenarios tested can be found in Table 3. The retention rates of 15 different commissioning sources were averaged and used to define ψ_{ary} for three of the scenarios. The average retention rate was then decreased by 4.5%, and this modified (i.e., "deflated") retention rate was used to define ψ_{ary} for the remaining three scenarios. The starting state of the system is also investigated. In Scenarios 1 and 2, the initial number of officers in each rank is equal to the target requirement for that rank. For example, if $\bar{S}_{a,2,y} = 200$, there are 200 captains in the starting state. These starting states are considered "on target." Also investigated were the cases where there are more junior officers in the starting state than senior officers ("bottom heavy") and where there are more senior officers in the starting state than junior officers ("top heavy").

Table 3. Scenarios

Scenario	Retention Rate	Starting State
1	Average	On Target
2	Deflated	On Target
3	Average	Bottom Heavy
4	Deflated	Bottom Heavy
5	Average	Top Heavy
6	Deflated	Top Heavy

For each scenario, the best θ is used to determine the performance of the ADP algorithm as measured by the mean cost and its 95% confidence interval. These results are compared to the 95% confidence interval for the benchmark policy. Refer to Table 4 for the results. In Scenarios 1 and 2, the ADP policy performs significantly better than the benchmark policy. In Scenarios 3 and 4, the benchmark policy performs significantly better than the ADP policy. In Scenarios 5 and 6, neither policy performs significantly better than the other. However, it should be noted that the confidence interval for the ADP policy is tighter in all scenarios, suggesting the ADP policy to be more dependable than the benchmark.

Table 4. LSTD Results: Quality of Solution with the Best θ

Scenario	Algorithm Parameters (N, M, η, a)	ADP 95% CI	Benchmark 95% CI
1	25, 5000, 10, 10	66.72 ± 0.52	121.38 ± 10.24
2	50, 1000, 100, 10	88.69 ± 0.29	158.29 ± 10.30
3	50, 1000, 10, 10	181.48 ± 0.29	127.88 ± 14.40
4	50, 1000, 10, 1	191.39 ± 0.50	163.44 ± 9.85
5	50, 1000, 100, 10	147.58 ± 0.21	148.22 ± 24.97
6	50, 5000, 100, 10	161.13 ± 0.25	177.35 ± 16.80

Table 5 refers to how robust the ADP algorithm is in each scenario. The five runs indicate the average cost of thirty iterations of five different θ-vectors. The best overall mean cost is reported with the corresponding algorithm parameters. It should be noted that the best mean found in Table 4 may have different algorithm features than those presented in Table 5. In fact, only Scenario 6 contains the overall minimum cost and the minimum averaged cost.

Table 5. LSTD Results: Robustness of Solutions

Scenario	Algorithm Parameters (N, M, η, a)	Run 1	Run 2	Run 3	Run 4	Run 5	Mean	Standard Deviation
1	50, 5000, 100, 10	77.36	66.80	66.91	66.83	67.02	68.99	4.68
2	50, 1000, 100, 10	92.79	92.81	92.47	92.71	88.69	91.90	1.80
3	25, 1000, 100, 1	202.85	203.38	181.96	181.74	181.68	190.32	11.68
4	50, 5000, 10, 10	211.90	191.48	191.55	211.88	211.68	203.70	11.12
5	50, 5000, 10, 1	156.17	155.63	155.51	155.88	155.65	155.77	0.26
6	50, 5000, 100, 10	161.60	161.47	161.13	161.33	161.61	161.43	0.20

Meta-Analysis on Algorithmic Features.

The four ADP LSTD algorithm parameters, policy improvement (N), policy evaluation (M), regularization (η), and harmonic smoothing (a), were tested for significance when determining the computational time it took to determine the θ-vectors for average retention, the computational time it took to determine the θ-vectors for deflated retention, and the cost for each of the six scenarios.

Table 6 summarizes the parameters that significantly impacted the cost per scenario. In Scenario 1, the harmonic smoothing a parameter and the interaction between the policy improvement, N, regularization, η, and harmonic smoothing parameters were significant at the 95% confidence interval. Refer to the table to see which parameters and interactions become significant at the 90% confidence interval. Scenario 2 did not contain any parameters of significance at the 95% confidence level. However, at the 90% confidence level, the interaction between the policy improvement and the harmonic smoothing parameters becomes significant. In Scenario 3, the interaction of the policy improvement, regularization, and harmonic smoothing parameters were significant. In Scenario 4, the interaction of the regularization and harmonic smoothing parameters was significant. In Scenario 5, the harmonic smoothing and the interaction between the policy improvement, regularization, and harmonic smoothing parameters were significant. Finally, in Scenario 6, the policy evaluation, M, the interaction between policy improvement and regularization, and the interaction between the regularization and harmonic smoothing parameters were significant.

Table 6. LSTD Results: Parameter P-Values

Term	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
N	0.798	0.194	0.832	0.762	0.294	0.134
M	0.164	0.309	†0.084	0.997	0.886	⋆0.033
η	0.198	0.102	0.182	0.831	0.162	0.584
a	⋆0.022	0.228	0.205	0.538	⋆0.006	0.278
$N \cdot M$	0.458	0.324	0.430	†0.092	0.380	0.138
$N \cdot \eta$	0.393	0.227	0.732	0.505	0.450	⋆0.002
$N \cdot a$	†0.063	†0.063	†0.059	0.872	†0.055	0.539
$M \cdot \eta$	0.788	0.197	0.687	0.831	0.766	0.120
$M \cdot a$	0.193	0.249	0.271	0.170	0.211	0.377
$\eta \cdot a$	0.804	0.250	0.732	⋆0.016	0.255	⋆0.001
$N \cdot M \cdot \eta$	†0.059	0.522	0.386	0.183	0.597	0.884
$N \cdot M \cdot a$	0.222	0.177	0.178	0.558	0.179	0.530
$N \cdot \eta \cdot a$	⋆0.017	0.171	⋆0.042	0.698	⋆0.023	0.916
$M \cdot \eta \cdot a$	0.426	0.155	0.346	0.673	0.731	0.704
$N \cdot M \cdot \eta \cdot a$	0.175	0.658	0.821	0.980	0.987	0.945
⋆	denotes statistical significance at the 95% confidence level					
†	denotes statistical significance at the 90% confidence level					

Scenario 1.

The ADP policy has the best performance in Scenario 1 when compared to both the benchmark and all other scenarios. It is likely that the promotion decisions in the ADP policy cause it to outperform the benchmark. The benchmark policy has deterministic promotion rates based only on rank, whereas the ADP policy utilizes an inner minimization problem (IMP), as found in Equation 12.

The most robust θ was with algorithmic features of $N = 50, M = 5,000, \eta = 100$, and $a = 10$. The best minimal average cost can be seen in Run 2. The standard deviation of the five runs is 4.68. To improve the policy further, exploring the algorithmic parameters N, a, and η, could increase policy performance and decrease the average cost found in Table 4.

Scenario 2.

In Scenario 2, the ADP policy still outperforms the benchmark policy. The cost, however, does increase as compared to Scenario 1. It is likely that the deflated retention rates impact the cost. The most robust θ was with algorithmic features of $N = 50, M = 1,000, \eta = 100$, and $a = 10$. None of the algorithmic parameters have statistical significance at the 95% confidence level, and only the interaction of N and η are statistically significant at the 90% confidence level. This indicates that the parameter levels chosen do not accurately capture the variance in the results. To improve the policy, the experiment should be run at different algorithmic feature levels.

Scenario 3.

Scenario 3 investigates the cases when there are more junior officers (i.e., second lieutenants, first lieutenants, and captains) in the starting state than senior officers (i.e., majors, lieutenants colonels, and colonels) and the average retention rate is utilized. The average minimal cost more than doubles from the on target/average case in Scenario 1. Also, the benchmark policy outperforms the ADP policy with statistical significance. The basis functions utilized for this ADP policy capture the absolute value of officers either over or under the targeted value. Exploring the number of officers in each rank or adding value to higher ranking officers in the basis function could improve overall ADP policy output.

The most robust θ was with algorithmic features of $N = 25, M = 1,000, \eta = 100$, and $a = 1$. The best minimal average cost can be seen in Run 5. The standard deviation of the five runs is 11.68. The standard deviation is larger than what was found in Scenarios 1 and 2. This is likely due to inadequate selection of the basis functions and the increased inherent variance of the retention random variables. While exploring the algorithmic parameters of N, M, a, and η could increase policy performance and decrease the average cost found in Table 4, it would likely not be enough to outperform the benchmark.

Scenario 4.

Similarly to Scenario 3, Scenario 4 investigates the starting state where there are more junior officers than senior officers. However, the deflated retention rate is now utilized. As with Scenario 2, the ADP policy performs worse with the deflated retention rates as compared to the average retention rates utilized in Scenario 3. Here, the benchmark policy outperforms the ADP policy with statistical significance.

The most robust θ was with algorithmic features of $N = 50, M = 5,000, \eta = 10$, and $a = 10$. The best minimal average cost can be seen in Run 2. The standard deviation of the five runs is 11.12. Again, this is higher than the standard deviations found in Scenarios 1 and 2. While exploring the algorithmic parameters of N, M, a, and η could increase policy performance and decrease the average cost found in Table 4, it would likely not be enough to outperform the benchmark. The selection of basis functions should be further explored.

Scenario 5.

Scenario 5 investigates the cases when there are more senior officers (i.e., majors, lieutenants colonels, and colonels) in the starting state than junior officers (i.e., second lieutenants, first lieutenants, and captains) and the average retention rate is utilized. The average minimal cost improves as compared to Scenarios 3 and 4, but it does not statistically outperform the benchmark policy. However, it should be noted that the variance in the ADP policy is much smaller than the variance in the benchmark policy, proving it to be a more reliable policy.

The most robust θ was with algorithmic features of $N = 50, M = 5,000, \eta = 10$, and $a = 1$. The best minimal average cost can be seen in Run 3. The standard deviation of the five runs is 0.26. This standard deviation is smaller than those seen in all previous scenarios. Exploring the algorithmic features of N, a, and η could increase policy performance and decrease the average cost found in Table 4. The basis functions suggested in Scenario 3 should be explored for this scenario to determine if it could perform over the benchmark policy with statistical significance.

Scenario 6.

Similarly to Scenario 5, Scenario 6 investigates the cases when there are more senior officers in the starting state than junior officers. However, the deflated retention rate is now utilized. As with Scenarios 2 and 4, the ADP policy performs worse with the deflated retention rates as compared to the average retention rates utilized in Scenario 5. The ADP policy does not outperform the benchmark policy with statistical significance. However, it should be noted that the variance in the ADP policy is much smaller than the variance in the benchmark policy, proving it to be a more reliable policy.

The most robust θ was with algorithmic features of $N = 50, M = 5,000, \eta = 100$, and $a = 10$. The best minimal average cost can be seen in Run 3. Scenario 6 has the best standard deviation across all scenarios with a value of 0.20. Exploring the algorithmic features of N, M, a, and η could increase the policy performance and decrease the average cost found in Table 4. The basis functions suggested in Scenario 3 should be explored for this scenario to determine if they could outperform the benchmark policy with statistical significance.

V. Conclusions

5.1 Conclusions

This thesis seeks to advance the work done by Bradshaw [8] on the United States Air Force Officer Manpower Planning Problem (AFO-MPP). The AFO-MPP models officer accessions, promotions, and the uncertainty of retention rates. The objective for the AFO-MPP is to identify the policy for accession and promotion decisions that minimizes the expected total discounted cost of maintaining the required number of officers in the manpower system over an infinite time horizon. The AFO-MPP is formulated as an infinite-horizon Markov decision problem (MDP), and a policy is found using approximate dynamic programming. A least-squares temporal differencing (LSTD) algorithm is employed to determine the best approximate policies. Six computational experiments are conducted with varying retention rates and officer manning starting conditions.

In Scenarios 1 and 2, the ADP policy outperforms the benchmark policy (i.e., current United States Air Force policy) with statistical significance. In Scenarios 3 and 4, the benchmark policy outperforms the ADP policy. In Scenarios 5 and 6, there is no statistical significance between the ADP and benchmark policies. However, the variance in the ADP policy is smaller, indicating a more reliable system as compared to the benchmark. The higher average costs found in Scenarios 3-6 indicate that the basis functions selected were not appropriate for these cases.

In general, it appears the algorithmic parameters chosen (i.e., policy improvement N, policy evaluation M, regularization η, and harmonic smoothing a), were appropriate. Each scenario, other than Scenario 2, has parameter and parameter interaction statistical significance at the 95% confidence level. At the 90% confidence level, all scenarios indicate parameter and parameter interaction statistical significance.

5.2 Future Work

The work of this thesis expands the preliminary work of the AFO-MPP. Future work should take this expansion of the AFO-MPP and refine the LSTD algorithm, through use of different basis functions, to improve performance for Scenarios 3-6. Further, applying an alternate ADP technique and algorithm, such as least squares policy evaluation (LSPE), could provide an improved ADP policy as compared to the benchmark.

A more precise indication of the benchmark policy could also lead to a better comparison against the ADP policy proposed. The benchmark policy was modeled after speaking to appropriate personnel analysts. However, the benchmark policy could have added complexity that was not captured due to unfamiliarity with the system.

This work only explored a single Air Force Specialty Code (AFSC) but is capable of looking at multiple AFSCs. It is not uncommon for officers to switch fields during their career. Exploring the cross-flow of officers between AFSCs would add realism and an additional decision for the ADP policy utilized.

Finally, the algorithmic parameters of the LSTD algorithm should be further investigated. The parameters used were found to be significant, but the values used may not have been the best choices. By exploring different parameter values, a better performing ADP policy could be found.

VI. Appendix

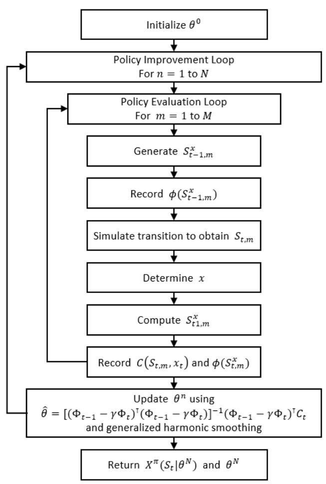
Approximate Dynamic Programming for the United States Air Force Officer Manpower Planning Problem

Capt Kimberly S. West, Advisor: Lt Col Matthew J. Robbins, Ph.D., Reader: Raymond R. Hill, Ph.D.

Background

- The United States Air Force (USAF) must attract and retain talented personnel
- This research seeks to improve policies regarding management of the commissioned officer corps to support mission readiness
- Management of the commissioned officer corps is a manpower planning problem
- Manpower planning problems, in general, determine the number of personnel needed to best meet current and future requirements
- One challenge the USAF faces is the closed nature of the military

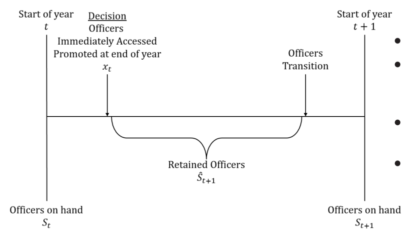
LSTD Algorithm



Contact Information

Lt Col Matthew J. Robbins, Ph.D.
Department of Operational Sciences, AFIT

Problem Description



- The decision is made at x_t
- Accession process is simplified because commissioning can happen at different points throughout the year
- The promotion process is simplified because officers can be promoted throughout the year
- Retention rates of officers with different AFSCO-rank-CYOS attributes may differ

Research Objectives

- Determine the manpower policy (i.e., hiring and promotion decisions as a function of the state of the personnel system) that minimizes expected total discounted cost of maintaining officers over an infinite horizon
- Formulate using Markov decision processes (MDP)
- Utilize approximate dynamic programming (ADP) algorithm to design, develop, and test a high quality manpower planning policy
- Compare ADP policy against current manpower practices
- Examine sensitivity of model parameters

Results and Conclusions

Scenario	Algorithm Parameters (N, M, η, a)	ADP 95% CI	Benchmark 95% CI
1	25, 5000, 10, 10	66.72 ± 0.52	121.38 ± 10.24
2	50, 1000, 100, 10	88.69 ± 0.29	158.29 ± 10.30
3	50, 1000, 10, 10	181.48 ± 0.29	127.88 ± 14.40
4	50, 1000, 10, 1	191.39 ± 0.50	163.44 ± 9.85
5	50, 1000, 100, 10	147.58 ± 0.21	148.22 ± 24.97
6	50, 5000, 100, 10	161.13 ± 0.25	177.35 ± 16.80

- Scenarios 1 and 2 the ADP policy outperforms the benchmark policy with statistical significance
- Scenarios 3 and 4, the benchmark policy outperforms the ADP policy with statistical difference
- Scenarios 5 and 6, there is no statistically significant difference between the ADP and benchmark policies
  - ⋆ Variance in the ADP policy is smaller, indicating a more reliable system as compared to the benchmark

- 2^4 full factorial design with five replicates
- All terms are free from aliasing
- Total of 80 runs
- ADP policy is created by calculating the θ coefficients for the basis functions from the implementation of the ADP algorithm
- Each run results in a different θ coefficient
- Use θ coefficients to conduct a simulation for the ADP policy and the benchmark policy over a 30 year horizon for 30 replications
- Common random numbers (CRN) utilized to reduce variance

MDP Model

- Decision Epochs: $\mathcal{T} = \{0, 1, 2, \ldots\}$
- States: $S_t = (S_{tary})_{(a,r,y)\in\mathcal{S}}$
  - AFSC of officer: $a \in \mathcal{S}^{AFSC}$, $\mathcal{S}^{AFSC} = \{1, 2, \ldots, A\}, A < \infty$
  - Rank of officer: $r \in \mathcal{S}^{rank}$ $\mathcal{S}^{rank} = \{1, 2, \ldots, 6\}$
  - CYOS of officer: $y \in \mathcal{S}^{CYOS}$ $\mathcal{S}^{CYOS} = \{1, 2, \ldots, 29\}$
- Decisions: $x_t = (x_t^{access}, x_t^{promote})$
  - Accession: $x_t^{access} = (x_{ta}^{access})_{a\in\mathcal{S}^{AFSC}}, \;\; x_{ta}^{access} \in \mathbb{N}^0$
  - Promotion: $x_t^{promote} = (x_{tary}^{promote})_{(a,r,y)\in\mathcal{S}^{promote}}$ $\mathcal{S}^{promote} = \cup_{a\in\mathcal{S}^{AFSC}}\Big\{(a,1,1),(a,2,3),$ $(a,3,9),(a,4,14),(a,5,19)\Big\} \subseteq \mathcal{S}$ $x_{tary}^{promote} \in \{0, 1, \ldots, S_{tary}\}$
- Transitions: Occur when
  - An officer is in a promotion window and promotes and retains or leaves the system
  - An officer is not in a promotion window and either stays or leaves the system
  - An officer is newly accessed to the system
  - An officer is in the promotion window but does not promote and retains or leaves the system
- Cost: In comparison to requirements
  - Over-manned
  - Under-manned
- Objective: $\min_{\pi\in\Pi} \mathbb{E}\Big\{\sum_{t=0}^{\infty}\gamma^t C(S_t, x_t)\Big\}$

Future Research

- Use of different basis functions, to improve performance for Scenarios 3-6
- Applying an alternate ADP technique and algorithm, such as least squares policy evaluation (LSPE)
- More precise indication of the benchmark policy
- Exploring the cross-flow of officers between AFSCs
- Algorithmic parameters of the LSTD algorithm should be further investigated

Bibliography

1. Ahner, Darryl K, & Parson, Carl R. 2015. Optimal multi-stage allocation of weapons to targets using adaptive dynamic programming. *Optimization Letters*, **9**(8), 1689–1701.

2. An, Lianjun, Jeng, Jun-Jang, Lee, Young M, & Ren, Changrui. 2007. Effective workforce lifecycle management via system dynamics modeling and simulation. *Pages 2187–2195 of: 2007 Winter Simulation Conference*. IEEE.

3. Banks, Jerry, Carson, John S., Nelson, Barry L., & Nicol, David M. 2010. *Discrete-event system simulation*. 5 edn. Prentice Hall, Upper Saddle River, NJ.

4. Bard, Jonathan F, Morton, David P, & Wang, Yong Min. 2007. Workforce planning at USPS mail processing and distribution centers using stochastic optimization. *Annals of Operations Research*, **155**(1), 51–78.

5. Bastian, Nathaniel D, McMurry, Pat, Fulton, Lawrence V, Griffin, Paul M, Cui, Shisheng, Hanson, Thor, & Srinivas, Sharan. 2015. The AMEDD uses goal programming to optimize workforce planning decisions. *Interfaces*, **45**(4), 305–324.

6. Bloch, Francis, & Cantala, David. 2013. Markovian assignment rules. *Social Choice and Welfare*, **40**(1), 1–25.

7. Blosch, Marcus, & Antony, Jiju. 1999. Experimental design and computer-based simulation: a case study with the Royal Navy. *Managing Service Quality: An International Journal*, **9**(5), 311–320.

8. Bradshaw, Amelia E. 2016. *United States Air Force officer manpower planning problem via approximate dynamic programming*. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.

9. Bradtke, Steven J, & Barto, Andrew G. 1996. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, **22**(1-3), 33–57.

10. Brook, Tom V. 2015. Defense secretary seeks overhal of military personnel system. *USA Today*, November. http://usat.ly/1j8bhf5.

11. Conley, Raymond E. 2006. *Maintaining the Balance Between Manpower, Skill Levels, and PERSTEMPO*. Vol. 492. RAND Corporation, Santa Monica, CA.

12. Conley, Raymond E, & Robbert, Albert A. 2009. *Air Force Officer Specialty Structure: Reviewing the Fundamentals*. RAND Corporation, Santa Monica, CA.

13. Davis, Michael T, Robbins, Matthew J, & Lunday, Brian J. 2016. Approximate dynamic programming for missile defense interceptor fire control. *European Journal of Operations Research*, **259**(3), 873–886.

14. Dimitriou, Vasileios, & Tsantas, Nikolaos. 2010. Evolution of a time dependent Markov model for training and recruitment decisions in manpower planning. *Linear Algebra and its Applications*, **433**(11), 1950–1972.

15. Dimitriou, Vasileios, Georgiou, Andreas, & Tsantas, Nikolaos. 2013. The multivariate non-homogeneous Markov manpower system in a departmental mobility framework. *European Journal of Operational Research*, **228**(1), 112–121.

16. Eoyang, Mieke, & Freeman, Ben. 2015 (May). *Why the U.S. must reform the Military Personnel System*. http://www.nationaldefensemagazine.org/archive/2015/May/Pages/WhytheUSMustReformTheMilitaryPersonnelSystem.aspx. Accessed: 2016-07-19.

17. Gans, Noah, & Zhou, Yong-Pin. 2002. Managing learning and turnover in employee staffing. *Operations Research*, **50**(6), 991–1006.

18. Gass, Saul I. 1991. Military manpower planning models. *Computers & Operations Research*, **18**(1), 65–73.

19. Gass, Saul I, Collins, Roger W, Meinhardt, Craig W, Lemon, Douglas M, & Gillette, Marcia D. 1988. OR Practice - The Army Manpower Long-Range Planning System. *Operations Research*, **36**(1), 5–17.

20. Gerontidis, Ioannis I. 1995. Periodicity of the profile process in Markov manpower systems. *European Journal of Operational Research*, **85**(3), 650–669.

21. Godfrey, Gregory A, & Powell, Warren B. 2001. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science*, **47**(8), 1101–1112.

22. Guerry, Marie-Anne. 2011. Hidden heterogeneity in manpower systems: A Markov-switching model approach. *European Journal of Operational Research*, **210**(1), 106–113.

23. Hoecherl, Joseph C, Robbins, Matthew J, Hill, Raymond R, & Ahner, Darryl K. 2016. Approximate dynamic programming algorithms for United States Air Force officer sustainment. *Pages 3075–3086 of: Winter Simulation Conference*. IEEE.

24. Khoong, Chan M. 1996. An integrated system framework and analysis methodology for manpower planning. *International Journal of Manpower*, **17**(1), 26–46.

25. Montgomery, Douglas C. 2008. *Design and Analysis of Experiments*. 8 edn. Hoboken, New Jersey: John Wiley & Sons.

26. Nicholls, Miles. 2009. The Use of Markov models as an aid to the evaluation, planning and benchmarking of Doctoral Programs. *Journal of the Operational Research Society*, **60**(9), 1183–1190.

27. Nilakantan, Kannan, & Raghavendra, B.G. 2005. Control aspects in proportionality Markov manpower systems. *Applied Mathematical Modelling*, **29**(1), 85–116.

28. Nirmala, S., & Jeeva, M. 2010. A dynamic programming approach to optimal manpower recruitment and promotion policies for the two grade system. *African Journal of Mathematics and Computer Science Research*, **3**(12), 297–301.

29. Onggo, Stephan, Pidd, Michael, Soopramanien, Didier, & Worthington, Dave. 2010. Simulation of career development in the European Commission. *Interfaces*, **40**(3), 184–195.

30. Powell, Warren B. 2011. *Approximate Dynamic Programming: Solving the curses of dimensionality*. 2 edn. Hoboken, New Jersey: John Wiley & Sons.

31. Powell, Warren B, & Van Roy, Benjamin. 2004. Approximate dynamic programming for high dimensional resource allocation problems. *Handbook of Learning and Approximate Dynamic Programming, IEEE Press, New York*, **8**.

32. Powell, Warren B, Bouzaiene-Ayari, Belgacem, Berger, Jean, Boukhtouta, Abdeslem, & George, Abraham P. 2011. The effect of robust decisions on the cost of uncertainty in military airlift operations. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **22**(1), 1.

33. Puterman, Martin L. 1994. *Markov Decision Processes, Discrete Stochastic Dynamic Programming*. Hoboken, New Jersey: John Wiley & Sons, Inc.

34. Rettke, Aaron J, Robbins, Matthew J, & Lunday, Brian J. 2016. Approximate dynamic programming for the dispatch of military medical evacuation assets. *European Journal of Operational Research*, **254**(3), 824–839.

35. Schofield, Jill A. 2015. *Non-rated Air Force line officer attrition rates using survival analysis*. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA.

36. Simão, Hugo P, George, Abraham, Powell, Warren B, Gifford, Ted, Nienow, John, & Day, Jeff. 2010. Approximate dynamic programming captures fleet operations for Schneider National. *Interfaces*, **40**(5), 342–352.

37. Škulj, Damjan, Vehovar, Vasja, & Štamfelj, Darko. 2008. The modelling of manpower by Markov chains-a case study of the Slovenian armed forces. *Informatica*, **32**(3).

38. Sohn, So Young, Chang, In Sang, & Moon, Tae Hee. 2007. Random effects Weibull regression model for occupational lifetime. *European Journal of Operational Research*, **179**(1), 124–131.

39. Song, Haiqing, & Huang, Huei-Chuen. 2008. A successive convex approximation method for multistage workforce capacity planning problem with turnover. *European Journal of Operational Research*, **188**(1), 29–48.

40. Tang, Qunhong, Wilson, George R, & Perevalov, Eugene. 2008. An approximation manpower planning model for after-sales field service support. *Computers & Operations Research*, **35**(11), 3479–3488.

41. Wang, Jun. 2005 (February). *A review of operations research applications in workforce planning and potential modeling of military training*. Tech. rept. Australian Government Department of Defence, Edinburgh, South Australia.

42. Weise, Thomas, Zapf, Michael, Chiong, Raymond, & Nebro, Antonio J. 2009. Why is optimization difficult? *Pages 1–50 of: Nature-Inspired Algorithms for Optimisation*. Springer.

43. White, Douglas J. 1985. Real applications of Markov decision processes. *Interfaces*, **15**(6), 73–83.

44. White, Douglas J. 1988. Further real applications of Markov decision processes. *Interfaces*, **18**(5), 55–61.

45. White, Douglas J. 1993. A survey of applications of Markov decision processes. *Journal of the Operational Research Society*, **44**(11), 1073–1096.

46. Workman, Patrick E. 2009. *Optimizing security force generation*. Ph.D. thesis, Monterey, California. Naval Postgraduate School.

47. Wu, Yan-Kuen. 2007. On the manpower allocation within matrix organization: a fuzzy linear programming approach. *European Journal of Operational Research*, **183**(1), 384–393.

48. Zhu, Xiaomei, & Sherali, Hanif D. 2009. Two-stage workforce planning under demand fluctuations and uncertainty. *Journal of the Operational Research Society*, **60**(1), 94–103.

Vita

Captain Kimberly S. West attended Newtown High School in Newtown, Connecticut and graduated in 2006. She earned her Bachelor of Science degree in Mathematics from Christopher Newport University in Newport News, Virginia in 2010. Kimberly was commissioned into the United States Air Force as a Second Lieutenant in April 2012.

Captain West's first assignment was to the 59th Test and Evaluation Squadron at Nellis AFB, Nevada. She was the lead analyst for the F-22 division. She participated in Acquisition I programs, as well as lead a Joint Urgent Operational Need (JUON), and was the liaison between F-22 developmental test (DT) and operational test (OT).

In August 2015, Kimberly entered the Air Force Institute of Technology's Graduate School of Engineering and Management at Wright-Patterson AFB, Ohio. Upon graduation, she will be assigned to Fort Lee in Petersburg, VA as an instructor for the Operations Research/Systems Analysis Military Applications Course (ORSA-MAC).

Permanent address: 2950 Hobson Way
Air Force Institute of Technology
Wright-Patterson AFB, OH 45433

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*
23–03–2017

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From — To)*
Aug 2015 – Mar 2017

4. TITLE AND SUBTITLE
Approximate Dynamic Programming
for the United State Air Force Officer
Manpower Planning Problem

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
West, Kimberly S., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-MS-17-M-162

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Gerald Diaz, Ph.D
Headquarters Air Force/A1
1550 W. Perimeter Rd, Rm 4710
Joint Base Andrews NAF Washington, MD 20762-5000
gerald.diaz.civ@mail.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
HAF/A1

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A.
Approved for Public Release; distribution unlimited.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The United States Air Force (USAF) makes officer accession and promotion decisions annually. Optimal manpower planning of the commissioned officer corps is vital. A manpower system that is neither over-manned nor under-manned is desirable as it is most cost effective. The Air Force Officer Manpower Planning Problem (AFO-MPP) is introduced, which models officer accessions, promotions, and the uncertainty in retention rates. The objective for the AFO-MPP is to identify the policy for accession and promotion decisions that minimizes expected total discounted cost of maintaining the required number of officers in the system over an infinite time horizon. The AFO-MPP is formulated as an infinite-horizon Markov decision problem, and a policy is found using approximate dynamic programming. A least-squares temporal differencing (LSTD) algorithm is employed to determine the best approximate policies possible. Six computational experiments are conducted with varying retention rates and officer manning starting conditions. The LSTD algorithm results are compared to the benchmark policy (e.g., currently practiced by the USAF). Results indicate that

15. SUBJECT TERMS
manpower planning, officer sustainment, workforce planning, Markov decision processes, MDP, approximate dynamic programming, ADP, approximate policy iteration, least squares temporal differences, LSTD

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	61	Lt Col Matthew J.D. Robbins, Ph.D.
					19b. TELEPHONE NUMBER *(include area code)* (937)255-3636, x4539; matthew.robbins@afit.edu

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18