Dynamic Surgery Management Under Uncertainty
(不确定性下的动态手术管理)

Wu Zhang

School of Economics and Business Administration, Chongqing University

Sep 4th, 2025

Paper Information

- **Journal:** European Journal of Operational Research (2023)
- **Authors:** E. Gökalp, N. Gülpnar, X. V. Doan
- **Article history:**Received 18 May 2021.Accepted 2 December 2022.Available 10 January 2023

Research Motivation

Importance of Surgery Management:

- Operating rooms are the "engine" of hospitals, generating 40% of total revenue
- Consume significant physical resources (beds, equipment) and human resources
- Existing management practices fail to achieve performance targets

Core Challenges:

1. **Surgery Duration Uncertainty** - Actual completion times are unpredictable
2. **Disruptive Events** - Equipment failures, random emergency arrivals
3. **Real-time Decision Requirements** - Need to dynamically update surgery allocation

Problem Severity:

- UK NHS: 1.1% of elective surgeries cancelled (79,470 cases/year)
- Capacity shortage is the main cancellation reason
- Long waiting times severely impact patient satisfaction

Research Objectives & Contributions

Research Objective:

Establish a real-time surgery management optimization model for multiple operating rooms, simultaneously minimizing:

- Number of uncompleted surgeries (weighted by urgency level)
- Patient waiting times (weighted by urgency level)

Main Contributions:

1. **Modeling Contribution:** First to establish stochastic dynamic programming model considering multiple uncertainties
2. **Algorithmic Contribution:** Develop Approximate Dynamic Programming (ADP) algorithm to solve curse of dimensionality
3. **Empirical Contribution:** Validate algorithm performance and practical applicability using real data

Literature Review & Research Positioning

Research Classification Framework:

Dimension	Categories
Modeling Methods	Integer Programming (IP), Two-stage Stochastic Programming (TS), Markov Decision Process (MDP)
Decision Types	Cancellation (C), Assignment (A), Patient Arrival (PA), Surgery Duration (SD))
Cost Components	Cancellation Cost (Cc), Waiting Time Cost (Wc)
Scheduling Approaches	Proactive Scheduling (PS), Reactive Scheduling (RS)

Research Gap:

- Most studies focus on elective surgery scheduling
- Lack of comprehensive real-time management research considering all relevant uncertainties
- Overlooked dynamic nature of real-time surgery management

Model Assumptions

1. **Patient Arrival Pattern**
   - All elective patients arrive at beginning of day
   - Emergency patients arrive randomly over time
2. **Surgery Execution Strategy**
   - No emergency surgeries are rejected
   - Patients can wait until end of shift
   - Operating rooms can handle multiple surgery types
3. **Decision Points**
   - Make allocation decisions whenever operating room becomes available
   - Consider patient priority and waiting time

State Space Definition

System State $\mathrm{S_t} = (n_t, \tau, p, a, \omega^t, \ell^t, f^t)$

Patient Information

- n_t: Total number of patients at time t
- τ: Surgery type of patient i
- p_i: Priority of patient i
- a_i: Arrival time of patient i

System Status

- ω_i^t: Waiting status of patient i (1=waiting, 0=in surgery)
- ℓ_r^t: Patient ID currently assigned to room r
- f_r^t: Completion status of room r (1=available, 0=occupied)

Action Space & State Transition

Action Space

- **Decision Variable:** x_r^t Patient assigned to operating room r
- **Feasible Action Set:** $X_r^t = \{i : 1 \leq i \leq n^t | \omega_i^t = 1, p_i \in F_r\} \cup \{0\}$
  where F_r is the set of surgery types that room r can perform

State Transition Mechanism

- **Execute Actions**: Update patient and room assignments
- **Emergency Arrivals**: Randomly generate new patients
- **Surgery Completions**: Update completion status based on duration distributions

Objective Function Design

Bi-objective Optimization Model

$$C(S_t) = \alpha \cdot c_d(S^t) + (1-\alpha)\sum_{t=1}^{t} c_w^t(S^t)$$

Objective 1: Minimize Uncompleted Surgeries (Weighted)

$$c_d(S^t) = \sum_{i=1}^{n_t} p_i \omega_i^t$$

Objective 2: Minimize Waiting Time (Weighted)

$$c_w^t(S^t) = \sum_{i=1}^{n^t} p_i (t - a_i)\omega_i^t$$

Weight Parameter α Interpretation:$\alpha = 0$: Pure waiting time minimization; $\alpha = 1$: Pure completion number maximization; $0 < \alpha < 1$: Balance between two objectives

Dynamic Programming Model

Bellman Equation

$$V_t(S_t) = \min_{x_r^t \in X_r^t} \left\{ (1-\alpha) \sum_{i=1}^{n_t} p_i(t+1-a_i)\omega_i^{t+1} + E[V_{t+1}(S_{t+1})] \right\}$$

Expected Value Calculation (Triple Uncertainty)

$$E[V_{t+1}(S_{t+1})] = E_{\eta^{t+1}} \left[E_{(\tilde{p},\tilde{r})} \left[E_{d_{x_r^t}} [V_{t+1}(S_{t+1})] \right] \right]$$

- Emergency patient arrival numbers uncertainty
- Emergency patient characteristics uncertainty
- Surgery duration uncertainty

Boundary Condition

$$V_T(S_T) = c_d(S_T)$$

Approximate Dynamic Programming Architecture

ADP Core Ideas:

- **Forward Simulation** - Start from initial state, simulate system evolution
- **Sampling Approximation** - Only visit states on simulation paths
- **Function Approximation** - Estimate values of unvisited states

Main Algorithm Components:

- **Lookup Table**: Store values of visited states
- **Double-pass Strategy**: Forward exploration + backward update
- **Basis Function Approximation**: Handle unseen states
- **Exploration-Exploitation Balance**: ε-greedy strategy

Detailed ADP Algorithm Flow

Step 0: Initialize lookup table and parameters

Step 1: Sample path generation (Monte Carlo simulation)

Step 2: State generation and action evaluation

- Random event generation (emergency arrivals, surgery completions)
- Action selection (exploration vs exploitation)
- Immediate cost calculation

Step 3: Value function approximation

- Lookup table query
- Basis function approximation (unvisited states)

Step 4: Backward pass update

Step 5: Convergence check

State Space Compression Technique

Patient Clustering Strategy

Clustering Key: $(arrival\_time, priority, surgery\_type)$

Before vs After Compression

Aspect	Before Compression	After Compression
State Representation	Individual patient indexing	Clustering by features
State Variables	$\omega_i^t \in \{0, 1\}$	$\omega_c^t \in \{0, 1, 2, \ldots\}$
Complexity	$O(2^{n_{max}})$	$O(n_{max}^m)$

Practical Effects

- Elective patients: Divided into 3 categories by surgery type
- Emergency patients: Each patient independent (different arrival times)
- Computational efficiency: 10-room problems solvable within 2 hours

Basis Function Approximation

Basis Function Design

$$V_t^b(S_t) = \psi \cdot [\alpha \sum_{i=1}^{n_t} \omega_i^t p_i + (1-\alpha) \sum_{i=1}^{n_t} \omega_i^t (T - a_i)]$$

Feature Extraction

- Sum of priorities of currently waiting patients
- Weighted sum of remaining time for waiting patients

Parameter Estimation

- Method: Linear regression based on lookup table data
- Objective: Minimize prediction error sum of squares
- Update: Dynamically adjusted as algorithm progresses

1. Motivation
2. Problem Modeling
3. Solution Approach
4. **Computational Experiments**
5. Conclusion

Experimental Design & Data Settings

Basic Experimental Parameters

Parameter	Setting	Reference
Planning Horizon	10 hours (20 half-hour periods)	Jung et al. (2017)
Number of Rooms	3	Basic experiment setting
Surgery Types	3 types	Jung et al. (2017)
Elective Patients	9 (3 per type)	Experiment setting

Probability Distribution Settings

- Surgery Duration: Log-normal distribution
  - Type 1: Mean 1.5, Std 1.7
  - Type 2: Mean 2.5, Std 1.9
  - Type 3: Mean 3.8, Std 1.9
- Emergency Arrivals: 12.5 % per period (1 patient), 4 % (2 patients)

Algorithm Performance Analysis

Small-scale Problem Validation

Problem Setting: 2 rooms, 6 periods, 5 electives, 2 surgery types

Metric	Exact DP	ADP Algorithm
Computation Time	3.2 hours	$<$1 minute
Action Match Rate	100%	95%
Optimality Gap	0%	$<$5%
State Visits	All	Partial (sampling)

Large-scale Problem Scalability

- 3 rooms: 17 minutes
- 5 rooms: $<$1 minute (with compression)
- 10 rooms: 2 hours (with compression)

Comparison with Heuristic Methods

Myopic Policy Design

Cost Measure: $c_m^t(i) = \alpha p_i + (1-\alpha) p_i (t - a_i)$ Decision Rule: Select patient with highest cost

Performance Comparison (1000 random scenarios)

Performance Metric	ADP Policy	Myopic Policy	Improvement
Waiting Time	89.19±0.79	96.7±0.54	7.8%
Completed Surgeries	8.81±0.07	7.67±0.09	14.9%

ADP Advantage Analysis

- Forward-looking: Considers future random event impacts
- Global Optimization: Avoids local optima of myopic decisions
- Robustness: Strong adaptability to uncertainty parameter changes

Parameter Sensitivity Analysis

Impact of Weight Parameter α

Value	Waiting Time	Completed Surgeries	Policy Characteristics
0.0	Minimum	Fewer	Favor quick surgeries
0.5	Balanced	Balanced	L-curve corner point
1.0	Higher	Maximum	Favor long surgeries

Surgery Type Assignment Patterns

- α increases:Short-duration surgery assignment rate decreases
- α increases:Long-duration surgery assignment rate increases
- Explanation: When optimizing completion numbers, tendency to handle more complex surgeries

Management Strategy Comparison

Scheduling Strategy Experiments

Flexible vs Block Scheduling

Strategy	Waiting Time	Completed Surgeries
Flexible (Base)	89.19±0.79	8.81±0.07
Block Scheduling	97.97±0.5	7.47±0.092

Mixed vs Separate Scheduling

Strategy	Waiting Time	Completed Surgeries
Mixed (Base)	40.13±1.06	6.47±0.04
Separate Scheduling	76.31±1.23	5.23±0.04

Management Insights:Higher flexibility leads to better performance - Maximize

1. Motivation
2. Problem Modeling
3. Solution Approach
4. Computational Experiments
5. **Conclusion**

Main Findings & Conclusions

Theoretical Contributions

1. Modeling Innovation: First to apply ADP to real-time surgery management
2. Algorithmic Innovation: Integrated approach of double-pass + state compression + basis function approximation
3. Complexity Breakthrough: Reduced exponential problem to polynomial solvability

Practical Value

1. Decision Support: Provides scientific real-time decision tools for hospital managers
2. Strategy Guidance: Validates significant advantages of dynamic strategies
3. Operational Optimization: Provides quantitative comparison basis for different hospital operation modes

Algorithm Performance Summary

- Computational Efficiency: 2 hours to solve 10-room real-scale problems
- Approximation Accuracy: 95% action matching, 5% optimality gap
- Practicality: Can handle complex constraints of real hospitals

Limitations & Future Research Directions

Current Limitations

1. Planning Scope: Currently limited to single-day planning
2. Parameter Sensitivity: Requires accurate estimation of uncertainty parameters
3. Constraint Considerations: Insufficient consideration of medical staff and equipment constraints

Future Research Directions

1. Multi-day Planning Extension:Consider cross-day impacts and resource constraints. Build rolling horizon optimization framework
2. Deep Learning Integration:Use neural networks to improve value function approximation. Adaptive learning of uncertainty distributions
3. Robustness Enhancement:Distributionally robust optimization methods.Performance guarantees under worst-case scenarios

Thanks!