הטכניון – מכון טכנולוגי לישראל
הפקולטה להנדסת חשמל

סימסטר אביב תשע״א

לימוד במערכות מורכבות (049004)

מרצה: פרופ׳ נחום שימקין, מאייר 653, טל. 4734 , דוא״ל shimkin@ee.technion.ac.il
שעות קבלה: ג׳ 15:30-17:30 (מומלץ תאום מראש)

דרישת קדם:
אותות אקראיים (או קורס מקביל בנושא תהליכים אקראיים).
קדם מומלץ: מערכות לומדות.

נושא הקורס: הקורס מציג מבוא מקיף לנושאים של **תכנות דינאמי מקורב** ו-**למידה באמצעות חיזוקים**, המהווים תחומי מחקר מרכזיים בתחומי הלמידה הממוחשבת ותכנון רב-שלבי. הבעייה המרכזית הינה בחירה אופטימלית של פעולות בבעיות החלטה רב שלביות, כאשר הדגש הוא על בעיות שבהן אופטימיזציה ישירה אינה אפשרית עקב סיבוכיות המערכת או חוסר מידע לגביה. המודל הבסיסי לבעיות מעין אלו הינו **תהליכי החלטה מרקוביים**. הפתרונות המוצעים מסתמכים על שילוב של טכניקות מתחום התכנון האופטימלי (ובפרט תכנות דינמי) עם שיטות למידה וקירוב. יידונו שימושים אופייניים בתחומי בקרת תנועה ברובוטיקה, משחקי לוח (כדוגמת שח), בקרת רשתות תקשורת, חקר ביצועים, לוגיסטיקה, ועוד.

הרכב הציון: 40% עבודות בית, 20% הצגת מאמר בכתה, 40% עבודה סמינריונית.

Basic Textbooks:

- D.P. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming,* Athena Scientific, 1996.
- R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction,* MIT Press, 1998.

Additional references:

- W.P Powell, *Approximate Dynamic Programming,* Wiley, 2007.
- K. Busonu, R. Babuska, B. Schytter abnd D. Ernts, *Reinforcement Learning and Dynamic Programming***,** CRC Press, 2010.
- C. Szepesvari, *Algorithms for Reinforcement Learning*, Morgan and Claypool Publishers, 2010.
- Papers from the current literature.

Course Outline:

1. **Introduction to Reinforcement Learning. (2 hours).**
2. **Dynamic Programming (6 hours):**
   Markov Decision Processes (MDPs)
   Finite and infinite horizon and horizon problems
   Value iteration, policy iteration, linear programming.
   On-line planing (preview)
3. **Basic Reinforcement Learning Algorithms (4 hours):**
   Monte Carlo methods
   TD(λ) and SARSA
   Q-learning
4. **Convergence analysis (4 hours)**
   The Stochastic Approximations algorithm
   Application to TD(0) and Q-learning
5. **Efficient Exploration (4 hours)**
   Mulit-armed bandits
   Efficient exploration algorithms for MDPs
6. **Value Function Approximations (4 hours)**
   Value function approximations: linear and nonlinear
   LSTD
   Actor-Critic algorithms.
7. **Policy Gradient Methods (2 hours)**
8. **Applications (2 hours, + reading)**