T.C.

AKDENİZ ÜNİVERSİTESİ

İNSANSI ROBOT YÜRÜYÜŞÜNÜN DERİN PEKİŞTİRMELİ ÖĞRENME TABANLI KONTROLÜ

Mustafa AYYILDIZ

FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

DOKTORA TEZİ

NİSAN 2026

ANTALYA

T.C.
AKDENİZ ÜNİVERSİTESİ

İNSANSI ROBOT YÜRÜYÜŞÜNÜN DERİN PEKİŞTİRMELİ ÖĞRENME TABANLI KONTROLÜ

Mustafa AYYILDIZ

FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

DOKTORA TEZİ

NİSAN 2026

ANTALYA

T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İNSANSI ROBOT YÜRÜYÜŞÜNÜN DERİN PEKİŞTİRMELİ ÖĞRENME TABANLI KONTROLÜ

Mustafa AYYILDIZ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

DOKTORA TEZİ

NİSAN 2026

Evrak Tarih ve Sayısı: 10.04.2026-904

T.C.
AKDENİZ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İNSANSI ROBOT YÜRÜYÜŞÜNÜN DERİN PEKİŞTİRMELİ ÖĞRENME TABANLI KONTROLÜ

Mustafa AYYILDIZ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

DOKTORA TEZİ

Bu tez 03/04/2026 tarihinde jüri tarafından Oybirliği ile kabul edilmiştir.

Prof. Dr. Övünç POLAT (Danışman)
Prof. Dr. Süleyman BİLGİN
Doç. Dr. Umut TİLKİ
Dr. Öğr. Üyesi Sıtkı GÜNER
Dr. Öğr. Üyesi Ufuk ÖZKAYA

Bu belge, güvenli elektronik İmza ile imzalanmıştır.
Evrak sorgulaması https://turkiye.gov.tr/ebd?eK=5543&eD=BSATTYRHT4&eS=904 adresinden yapılabilir. (PIN:75592)

ÖZET

İNSANSI ROBOT YÜRÜYÜŞÜNÜN DERİN PEKİŞTİRMELİ ÖĞRENME TABANLI KONTROLÜ

Mustafa AYYILDIZ

Doktora Tezi, Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Övünç POLAT

Nisan 2026; 104 sayfa

Bu tez çalışmasında, sagittal düzlem ağırlıklı harekete sahip üç boyutlu bir insansı robot modelinde derin pekiştirmeli öğrenme tabanlı yürüyüş denetimi incelenmiş ve özellikle kol salınım stratejilerinin yürüyüş performansı üzerindeki etkileri evrimsel strateji destekli hibrit bir öğrenme yapısı çerçevesinde değerlendirilmiştir. Çalışmada Simulink/Simscape Multibody ortamında oluşturulan, torkla sürülen ve toplam sekiz aktif serbestlik derecesine sahip bir insansı model kullanılmıştır. Bu serbestlik derecelerinin altısı alt ekstremitelere, ikisi ise omuz eklemlerine atanmıştır. Böylece üç boyutlu fiziksel yapı korunurken, öğrenme sürecini yönetilebilir kılacak ölçüde sadeleştirilmiş fakat biyomekanik açıdan anlamlı bir modelleme altyapısı oluşturulmuştur.

İzlenen yöntem iki aşamalı olarak yapılandırılmıştır. İlk aşamada ES-SAC, ES-TD3 ve ES-DDPG ajanları ile bunların saf karşılıkları olan SAC, TD3 ve DDPG algoritmaları aynı problem kurulumu altında karşılaştırılmıştır. Değerlendirmelerde normalize eğri altı alan, son dönem ödülü, en iyi pencere ödülü, hedefe ulaşma bölüm sayısı ve parkur tamamlama süresi gibi ölçütler kullanılmıştır. Elde edilen bulgular, evrimsel strateji bileşeninin keşif yeteneğini güçlendirdiğini, öğrenme sürecinin dayanıklılığını artırdığını ve genel performansa olumlu katkı sağladığını göstermiştir. Bu aşamada ES-SAC yaklaşımı, genel öğrenme başarımı ve kararlılık bakımından en güçlü yöntem olarak belirlenmiştir.

İkinci aşamada, en başarılı hibrit ajan olan ES-SAC kullanılarak üç farklı kol stratejisi düz ve engebeli zemin koşullarında ayrıntılı biçimde incelenmiştir. Bu stratejiler, kolların tamamen sabit tutulduğu sabit kol durumu, kolların aynı taraftaki bacaklarla aynı fazda salındığı anti-normal durum ve insan yürüyüşüne benzer karşı fazlı kol-bacak koordinasyonunu temsil eden normal durumdur. Bu aşamada, yürüyüş kararlılığı, ileri yönlü hareket sürekliliği, gövde dengesi ve enerjiyle ilişkili davranışları birlikte teşvik eden özel bir ödül fonksiyonu tanımlanmış, kol salınım stratejileri de omuz eklemlerinin ilgili faz ilişkilerine göre modellenmiştir. Böylece kol hareketi yalnızca kinematik bir tercih olarak değil, öğrenme sürecini ve yürüme davranışını doğrudan etkileyen bir denetim bileşeni olarak ele alınmıştır. Her koşul için beş bağımsız rastgele tohum ile eğitim yapılmış, ödül eğrileri ve yürüyüş kalitesi skoru ile birlikte ileri hız, yanal sapma, düşey salınım, toplam güç, toplam tork, birim hız başına güç ve birim hız başına tork gibi fiziksel ve davranışsal göstergeler üzerinden çok ölçütlü bir değerlendirme gerçekleştirilmiştir. Ayrıca eklem davranışlarının çevrimsel yapısını incelemek amacıyla olay hizalamalı limit çevrim faz portreleri kullanılmış, böylece analiz

yalnızca ödül temelli bir karşılaştırma ile sınırlı bırakılmayarak yürüyüşün yapısal niteliğini yansıtan bir değerlendirme çerçevesi oluşturulmuştur.

Elde edilen bulgular, kol salınım biçimlerinin denetim başarımı üzerindeki etkisini ortaya koymanın yanı sıra insan yürüyüşünden elde edilen deneysel çıktılarla da karşılaştırılmıştır. Bu karşılaştırmalar, insan yürüyüşünde doğal olarak gözlenen karşı fazlı kol-bacak koordinasyonunun ritmik süreklilik, dengeleyici gövde davranışı ve enerji ekonomisi bakımından bu çalışmada elde edilen robot yürüyüş bulgularıyla genel olarak uyumlu olduğunu göstermiştir. Bununla birlikte, robot deneylerinde anti-normal kol stratejisinin bazı görev odaklı ölçütlerde, özellikle ileri hız ve birim hız başına tork gibi göstergelerde avantaj sağlayabildiği gözlenmiştir. Buna karşılık normal kol salınımı, öğrenme kararlılığı, ödül düzeyi ve yürüyüş kalitesi bakımından daha dengeli ve daha istikrarlı bir yapı ortaya koymuştur. Sabit kol konfigürasyonu ise hem öğrenme hem de yürüyüş niteliği bakımından diğer stratejilerin gerisinde kalmıştır.

Sonuç olarak bu tez, hibrit ES-SAC denetimi altında kol hareketinin insansı robot yürüyüşünde yalnızca biyomekanik görünümü değil, aynı zamanda öğrenme verimliliğini, kararlılığı, hız özelliklerini, izleme başarımını ve enerjiyle ilişkili ölçütler arasındaki ödünleşimleri de belirgin biçimde şekillendirdiğini ortaya koymaktadır. Çalışma, insan yürüyüşünden esinlenen doğal kol salınımının dengeleyici ve ekonomik bir çözüm sunduğunu, ancak robotik denetim bağlamında alternatif kol stratejilerinin de belirli performans hedefleri altında işlevsel üstünlükler üretebildiğini göstermektedir. Bu yönüyle tez, kol salınımını ikincil bir hareket unsuru olarak değil, insansı lokomosyonda işlevsel ve tasarımsal öneme sahip bir denetim değişkeni olarak konumlandırmaktadır.

ANAHTAR KELİMELER: Derin Pekiştirmeli Öğrenme, Evrimsel Strateji, İnsansı Robot Yürüyüşü, Kol Salınımı, Yumuşak Aktör-Eleştirmen

JÜRİ: Prof. Dr. Övünç POLAT

Prof. Dr. Süleyman BİLGİN

Doç. Dr. Umut TİLKİ

Dr. Öğr. Üyesi Sıtkı GÜNER

Dr. Öğr. Üyesi Ufuk ÖZKAYA

ABSTRACT

DEEP REINFORCEMENT LEARNING-BASED CONTROL OF HUMANOID ROBOT WALKING

Mustafa AYYILDIZ

PhD Thesis in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Övünç POLAT

April 2026; 104 pages

In this dissertation, deep reinforcement learning based locomotion control is investigated for a three-dimensional humanoid robot model whose motion is predominantly governed by the sagittal plane, and the effects of arm swing strategies on walking performance are evaluated within an evolutionary strategy assisted hybrid learning framework. The study employs a torque driven humanoid model developed in the Simulink/Simscape Multibody environment with a total of eight active degrees of freedom. Six of these active degrees of freedom are assigned to the lower limbs, while the remaining two correspond to the shoulder joints. In this way, the three-dimensional physical structure is preserved while a simplified, yet biomechanically meaningful modelling framework is established to keep the learning process tractable.

The methodology is organized in two stages. In the first stage, ES-SAC, ES-TD3, and ES-DDPG agents are compared with their vanilla counterparts SAC, TD3, and DDPG under the same problem formulation. The evaluation includes normalized area under the learning curve, final period reward, best window reward, episodes to target, and track completion time. The findings show that the evolutionary strategy component strengthens exploration capability, improves the robustness of the learning process, and contributes positively to overall performance. At this stage, ES-SAC is identified as the strongest method in terms of overall learning performance and stability.

In the second stage, the best performing hybrid agent, ES-SAC, is used to investigate three arm swing strategies under both flat and rough terrain conditions. These strategies are the fixed-arm condition, in which both arms are kept locked, the anti-normal condition, in which the arms swing in phase with the ipsilateral legs, and the normal condition, which represents the human-like contralateral arm-leg coordination pattern. At this stage, a task specific reward function is designed to jointly promote walking stability, forward progression continuity, trunk balance, and energy related behaviours, while the arm swing strategies are modelled according to the phase relationships imposed at the shoulder joints. Thus, arm motion is treated not merely as a kinematic preference but as a control component that directly shapes both the learning process and the resulting locomotion behaviour. For each condition, training is conducted with five independent random seeds, and a multi-criteria evaluation is performed using reward curves, walking quality score, forward velocity, lateral deviation, vertical oscillation, total power, total torque, power per unit speed, and torque per unit speed. In addition, event aligned limit cycle phase portraits are employed to examine the cyclic structure of joint behaviour,

allowing the analysis to go beyond reward-based comparison and capture the structural characteristics of locomotion.

The findings reveal not only how different arm swing patterns affect control performance but also how these outcomes relate to experimental findings reported for normal human walking. These comparisons indicate that the naturally occurring contralateral arm-leg coordination observed in human gait is generally consistent with the robot results in terms of rhythmic continuity, stabilizing whole-body behaviour, and energy economy. However, the anti-normal arm strategy is found to provide advantages in certain task-oriented metrics, particularly forward speed and torque per unit speed. In contrast, the normal arm swing strategy yields a more balanced and stable profile in terms of learning stability, reward level, and walking quality. The fixed-arm configuration performs worse than the other strategies in both learning and locomotion quality.

Overall, this dissertation demonstrates that under hybrid ES-SAC control, arm motion shapes not only the biomechanical appearance of humanoid walking but also learning efficiency, stability, speed related behaviour, tracking performance, and the trade-offs among energy related metrics. The study shows that natural arm swing inspired by human gait offers a stabilizing and economical solution, while alternative arm strategies may still produce functional advantages under specific performance objectives in robotic control. In this respect, the dissertation positions arm swing not as a secondary motion feature but as a functionally significant control variable in humanoid locomotion.

KEYWORDS: Arm Swing, Deep Reinforcement Learning, Evolutionary Strategy, Humanoid Robot Locomotion, Soft Actor-Critic

COMMITTEE: Prof. Dr. Övünç POLAT

Prof. Dr. Süleyman BİLGİN

Assoc. Prof. Dr. Umut TİLKİ

Asst. Prof. Dr. Sıtkı GÜNER

Asst. Prof. Dr. Ufuk ÖZKAYA

ÖNSÖZ

Günümüzde insansı robot teknolojileri yalnızca laboratuvar ortamlarında geliştirilen deneysel sistemler olmaktan çıkmakta, endüstriyel alanlardan sağlık uygulamalarına, insan destekli hizmetlerden gündelik yaşam ortamlarına kadar uzanan geniş bir kullanım alanına doğru ilerlemektedir. İnsanla aynı fiziksel çevreyi paylaşabilen, onun hareket yapısına uyum sağlayabilen ve karmaşık görevleri öğrenebilen robot sistemlerinin yakın gelecekte yaşamın daha görünür bir parçası hâline geleceği açıkça görülmektedir. Bu nedenle insansı robotların yürüyüş yeteneğinin güvenilir, dengeli, enerji bakımından etkin ve öğrenilebilir biçimde geliştirilmesi, yalnızca akademik açıdan değil, toplumsal ve teknolojik açıdan da önemli bir araştırma alanı niteliği taşımaktadır. Bu tez, insansı robot yürüyüş denetimi ve kol salınım stratejilerinin söz konusu bağlamdaki rolünü daha derinlikli biçimde ortaya koymayı amaçlamaktadır.

Bu tez kapsamında, evrimsel strateji destekli yumuşak aktör-eleştirmen tabanlı bir öğrenme yaklaşımı ile insansı robot yürüyüş denetimi ele alınmış, kol salınım stratejilerinin yürüyüş kararlılığı, öğrenme başarımı ve enerjiyle ilişkili davranışlar üzerindeki etkileri incelenmiştir. Bu süreç, benim için yalnızca akademik bir çalışma değil, aynı zamanda sabır, disiplin ve süreklilik gerektiren önemli bir gelişim yolculuğu olmuştur.

Doktora eğitimim ve tez çalışmam boyunca bilgi, deneyim ve akademik birikimiyle bana yol gösteren, çalışmanın her aşamasında değerli görüş ve katkılarını esirgemeyen danışman hocam Prof. Dr. Övünç Polat'a en içten teşekkürlerimi sunarım. Akademik bakış açımın gelişmesinde ve bu çalışmanın şekillenmesinde kendisinin katkısı benim için son derece kıymetlidir.

Akademik yolculuğumun önceki aşamalarında bilgi, yönlendirme ve katkılarıyla gelişimime destek olan yüksek lisans danışman hocam Doç. Dr. Umut Tilki'ye de ayrıca içten teşekkürlerimi sunarım.

Hayatım boyunca her zaman yanımda olan, emekleri, destekleri ve bana duydukları güvenle en büyük gücüm olan aileme gönülden teşekkür ederim. Başta annem Hayriye Ayyıldız ve babam Sedat Ayyıldız olmak üzere, her zaman desteğini hissettiğim abim Dr. Fevzi Ayyıldız'a ve sevgili kız kardeşim Irmak Ayyıldız'a içten teşekkürlerimi sunarım. Bu uzun ve yoğun süreçte onların manevi desteği, sabrı ve varlığı benim için çok değerli olmuştur.

Ayrıca bu çalışmanın doğrudan ya da dolaylı biçimde oluşmasına katkı sağlayan tüm hocalarıma, arkadaşlarıma ve emeği geçen herkese teşekkür ederim.

Bu tezin, insansı robotların öğrenme tabanlı denetimi alanında çalışan araştırmacılar için yararlı bir kaynak olması en büyük temennimdir. Bunun yanında, bu çalışmanın benim için robotik alanında daha ileri araştırmaların sağlam bir temeli olmasını ve gelecekte araştırmalarımı yalnızca robotun fiziksel davranışlarıyla sınırlamayan, insan yaşamına daha doğal, daha uyumlu ve daha anlamlı biçimde eşlik edebilen gelişmiş robotik sistemlere taşımasını diliyorum.

İÇİNDEKİLER

ÖZET i
ABSTRACT iii
ÖNSÖZ v
AKADEMİK BEYAN vii
SİMGELER VE KISALTMALAR viii
ŞEKİLLER DİZİNİ xii
ÇİZELGELER DİZİNİ xiv
1. GİRİŞ 1
2. KAYNAK TARAMASI 8
3. MATERYAL VE METOT 14
3.1. İnsansı Robot Modeli 14
3.2. İnsansı Robot Bacağının Kinematik ve Dinamik Modeli 17
3.2.1. Düzlemsel üç serbestlik dereceli bacağın kinematik modeli 18
3.2.2. Düzlemsel bacak için hız kinematiği (Jakobiyen) ve statik modellemesi 26
3.2.3. Düzlemsel bacağın Lagrange tabanlı dinamik modellemesi 29
3.2.4. Düzlemsel bacakta kinematik kısıtlar 37
3.3. Evrimsel Strateji Destekli Yumuşak Aktör-Eleştirmen Modeli 40
3.3.1. Evrimsel strateji algoritmasının mimarisi 40
3.3.2. Yumuşak aktör-eleştirmen algoritmasının mimarisi 43
3.4. Yumuşak Aktör-Eleştirmen Ödül Fonksiyonu 52
3.5. Yumuşak Aktör-Eleştirmen Hiperparametreleri 56
3.6. SAC, TD3 ve DDPG Algoritmalarının Karşılaştırılması 58
3.7. ES-SAC Hibrit Yapısının Entegrasyonu 61
4. BULGULAR VE TARTIŞMA 65
4.1. Düz Zeminde Denetleyici Performanslarının Karşılaştırılması 65
4.2. Kol Salınım Stratejilerinin Düz ve Engebeli Zeminde İncelenmesi 73
5. SONUÇLAR 94
6. KAYNAKLAR 97
ÖZGEÇMİŞ

AKADEMİK BEYAN

Doktora Tezi olarak sunduğum “İnsansı Robot Yürüyüşünün Derin Pekiştirmeli Öğrenme Tabanlı Kontrolü” adlı bu çalışmanın, akademik kurallar ve etik değerlere uygun olarak yazıldığını belirtir, bu tez çalışmasında bana ait olmayan tüm bilgilerin kaynağını gösterdiğimi beyan ederim.

03/04/2026

Mustafa AYYILDIZ

SİMGELER VE KISALTMALAR

Simgeler

a_t : t anındaki eylem vektörü

c : $\cos\theta_2$ değeri

C : Coriolis ve merkezkaç terimlerini içeren matris veya vektör

D : Atalet matrisi

d : Kalça ile ayak bileği arasındaki doğrusal uzaklık

F : Uç noktaya etkiyen kuvvet vektörü

f_i : Kinematik kısıt fonksiyonları

G : Yerçekimi terimleri vektörü

g : Yerçekimi ivmesi

J : Jakobiyen matrisi

K : Değerlendirme sayısı

l_1 : Kalça ile diz arasındaki bağlantı uzunluğu

l_2 : Diz ile ayak bileği arasındaki bağlantı uzunluğu

l_3 : Ayak bileği ile ayak ucu arasındaki bağlantı uzunluğu

L : Lagrange fonksiyonu (L=T-V)

M : Mini yığın büyüklüğü veya örnek sayısı

n : Örnek sayısı

P : Güç veya konum noktası gösterimi, bağlama göre kullanılmıştır

|P| : Ortalama mutlak toplam eklem gücü

P_d : İstenen ayak ucu konumu

q : Genelleştirilmiş eklem koordinatları vektörü

$\dot{q}$: Eklem hızları vektörü

$\ddot{q}$: Eklem ivmeleri vektörü

Q : Durum-eylem değer fonksiyonu

Q_0 : Yürüme kalitesi skoru

r : Anlık ödül

R : Bölüm sonu toplam ödül veya getiri

T : Kinetik enerji

t : Zaman

V : Potansiyel enerji

V_x : Yanal hız

V_y : İleri hız

x : Yanal konum

|X| : Ortalama mutlak yanal konum

y : İleri yön konumu

z : Düşey konum

z_d : İstenen düşey konum bileşeni

y_d : İstenen ileri yön konum bileşeni

α : Entropi sıcaklık katsayısı veya geometrik açı, bağlama göre kullanılmıştır

β : Ters kinematikte kullanılan geometrik açı

ϵ : Rastgele gürültü vektörü

γ : İskonto katsayısı

λ : Lagrange çarpanları vektörü

μ_θ : Deterministik politika çıktısı veya ortalama eylem fonksiyonu

π_θ : Parametreleri θ olan stokastik politika

σ : Standart sapma veya mutasyon ölçeği

τ : Eklem torku veya hedef ağ güncelleme katsayısı, bağlama göre kullanılmıştır

|τ| : Ortalama mutlak toplam eklem torku

ϕ_d : İstenen ayak ucu yönelimi

ϕ_k : k. eleştirmen ağının parametre vektörü

φ : Uç nokta yönelimi

θ : Politika parametre vektörü veya genel açı gösterimi, bağlama göre kullanılmıştır

θ_1 : Kalça eklem açısı

θ_2 : Diz eklem açısı

θ_3 : Ayak bileği eklem açısı

w : Açısal hız

Bu tez çalışmasında ondalık ayracı olarak nokta (".") kullanılmıştır.

Kısaltmalar

AUC : Eğri altında kalan alan

nAUC : Normalize eğri altı alan

CEM : Çapraz Entropi Yöntemi

CoM : Kütle merkezi

CoT : Taşıma maliyeti

DDPG : Derin Deterministik Politika Gradyanı

DNN : Derin sinir ağı

DoF : Serbestlik derecesi

DRL : Derin pekiştirmeli öğrenme

ERL : Evrimsel pekiştirmeli öğrenme

ES : Evrimsel strateji

ES-DDPG : Evrimsel strateji destekli derin deterministik politika gradyanı

ES-RL : Evrimsel strateji destekli pekiştirmeli öğrenme

ES-SAC : Evrimsel strateji destekli yumuşak aktör-eleştirmen

ES-TD3 : Evrimsel strateji destekli ikiz gecikmeli derin deterministik politika gradyanı

IQR	: Çeyrekler arası aralık
LIPM	: Doğrusal ters sarkaç modeli
RL	: Pekiştirmeli öğrenme
SAC	: Yumuşak aktör-eleştirmen
TD3	: İkiz gecikmeli derin deterministik politika gradyanı
ZMP	: Sıfır moment noktası

ŞEKİLLER DİZİNİ

Şekil 1.1. Çalışmada izlenen iki aşamalı deneysel yöntemin genel akış şeması 6

Şekil 2.1. Çalışmada kullanılan insansı robot modelinin yandan görünüşü 9

Şekil 2.2. Çalışmada kullanılan indirgenmiş serbestlik dereceli insansı robot modelinin önden görünüşü 10

Şekil 2.3. a) İnsan yürüyüşünün sagittal düzlemde ardışık uzuv konumları; **b)** İnsan yürüyüşünün frontal düzlemde ardışık uzuv konumları 11

Şekil 3.1. İnsansı robot modelinin önden görünüşte eklem yerleşimleri ve koordinat eksenleri 15

Şekil 3.2. İnsansı robot modelinin perspektif görünüşte eklem yerleşimleri ve koordinat eksenleri 16

Şekil 3.3. a) Üst ekstremitede omuz ekleminin açısal limitleri; **b)** Alt ekstremitede kalça, diz ve ayak bileği eklemlerinin açısal limitleri 17

Şekil 3.4. Sağ bacağın sagittal düzlemdeki düzlemsel 3R kinematik modeli, eklem açıları ve bağlantı uzunlukları 18

Şekil 3.5. Sağ bacağın sagittal düzlemde ters kinematik çözümünde kullanılan geometrik açı tanımları ve bağlantı uzunlukları 24

Şekil 3.6. SAC ajanının genel mimarisi ve aktör-eleştirmen bileşenleri (Zhang vd. 2024) 44

Şekil 3.7. Çalışmada kullanılan aktör ağının katman mimarisi 48

Şekil 3.8. Çalışmada kullanılan eleştirmen ağının katman mimarisi 50

Şekil 3.9. ES-SAC hibrit algoritmasının genel mimarisi ve eğitim sürecindeki bileşen etkileşimleri (Pourchot ve Sigaud, 2019) 62

Şekil 4.1. 50 m'lik düz parkur üzerinde hareket eden insansı robot 65

Şekil 4.2. Düz zeminde serbest kol salınımı koşulunda eğitilen ajanlara ait öğrenme eğrileri (ortalama ± standart sapma) 69

Şekil 4.3. Düz zeminde serbest kol salınımı koşulunda eğitilen ajanlara ait $Q0$ değerlerinin bölüm sayısına göre değişimi 70

Şekil 4.4. ES-SAC ve ES-TD3 ajanları için yürüyüş yolu boyunca yanal sapmanın karşılaştırılması 71

Şekil 4.5. ES-SAC ve ES-TD3 ajanları için eklem bazında ortalama mutlak güç değerlerinin karşılaştırılması 72

Şekil 4.6. a) Sabit kol konfigürasyonu, her iki kolun tamamen kilitli olduğu ve kol salınımına izin verilmediği durum; **b)** anti-normal kol konfigürasyonu, kolların ipsilateral, yani aynı taraftaki bacak ile aynı fazda salındığı durum; **c)** normal kol konfigürasyonu, insan benzeri karşı fazlı kol-bacak koordinasyonunun sağlandığı durum 73

Şekil 4.7. Tüm jenerasyon boyunca ortalama ödülün medyan ve IQR ile değişimi..... 77

Şekil 4.8. Kol salınım stratejilerine göre ortalama ödül tohum medyanlarının kutu grafiği 78

Şekil 4.9. Kol salınım stratejilerine göre $Q0$ tohum medyanlarının kutu grafiği 79

Şekil 4.10. Kol salınım stratejilerine göre ileri hızın normalize zaman boyunca değişimi 80

Şekil 4.11. Kol salınım stratejilerine göre yanal hızın normalize zaman boyunca değişimi 81

Şekil 4.12. Kol salınım stratejilerine göre düşey sapmanın normalize zaman boyunca değişimi 82

Şekil 4.13. Kol salınım stratejilerine göre yanal sapmanın normalize zaman boyunca değişimi 83

Şekil 4.14. Kol salınım stratejilerine göre toplam eklem gücünün normalize zaman boyunca değişimi 84

Şekil 4.15. Kol salınım stratejilerine göre toplam eklem torkunun normalize zaman boyunca değişimi 85

Şekil 4.16. Kol salınım stratejilerine göre adım başına tümleşik mekanik iş dağılımı... 86

Şekil 4.17. Kol salınım stratejilerine göre mesafe başına tümleşik mekanik iş dağılımı 87

Şekil 4.18. Kol salınım stratejilerine göre mekanik taşıma maliyeti dağılımı..... 88

Şekil 4.19. Üç kol salınım konfigürasyonu için olay hizalamalı limit-çevrim faz portreleri: **a)** sağ omuz; **b)** sağ kalça; **c)** sağ diz; **d)** sağ ayak bileği; **e)** sol omuz; **f)** sol kalça; **g)** sol diz; **h)** sol ayak bileği; omuz eklemlerinde normal ve anti-normal, alt ekstremite eklemlerinde ise sabit kol, normal ve anti-normal konfigürasyonlarına ait yoğunluk konturları gösterilmektedir..... 89

Şekil 4.20. Engebeli zeminde insansı robot yürüyüşü 90

ÇİZELGELER DİZİNİ

Çizelge 3.1. Ödül/ceza terimlerinin değer aralıkları 53

Çizelge 3.2. Ödül ve ceza terimleri için ağırlık değerleri 53

Çizelge 3.3. Üç kol-salınım koşulu için ödül fonksiyonu tanımları 55

Çizelge 3.4. Ajan hiperparametre değerleri 56

Çizelge 3.5. SAC, TD3 ve DDPG algoritmalarının ve bunların ES tabanlı hibrit sürümleri olan ES-SAC, ES-TD3 ve ES-DDPG'nin politika yapısı, keşif mekanizması, hedef değer ve kayıp fonksiyonları açısından karşılaştırılması 60

Çizelge 4.1. Hibrit ve saf RL ajanlarının karşılaştırmalı öğrenme ve görev performansları 68

Çizelge 4.2. Üç yürüyüş konfigürasyonunun, normal, anti-normal ve sabit kol, aktif omuz serbestlik dereceleri, omuz tork aktüasyonu, distal kol eklemlerinin kilitlenme durumu, kol hareket düzlemi kısıtı, uygulanan kol-bacak faz kuralı ve buna karşılık gelen eylem uzayı boyutu açısından tanımları 74

Çizelge 4.3. Ortalama ödül ve yürüme kalitesi skoru ($Q0$) üzerindeki etkiler; değerler medyan [IQR], n = 5 olarak verilmiştir 75

Çizelge 4.4. Düz zemin üzerinde kararlı durum penceresinde hesaplanan yürüyüş metrikleri. Koordinatlar: ileri yön y, yanal yön x, düşey yön z. Vy: ileri hız (m/s); $|Vx|$: ortalama mutlak yanal hız (m/s); dz: ortalama düşey yer değiştirme (m); |X|: ortalama mutlak yanal konum (m); |P|: ortalama mutlak toplam eklem gücü (W); |P|/Vy: birim hız başına güç (J/m); $|\tau|$: ortalama mutlak toplam eklem torku (N·m); $|\tau|/Vy$: birim hız başına tork (N·m·s/m) 76

Çizelge 4.5. Engebeli zemin üzerinde kararlı durum penceresinde hesaplanan yürüyüş metrikleri. Koordinatlar: ileri yön y, yanal yön x, düşey yön z. Vy: ileri hız (m/s); $|Vx|$: ortalama mutlak yanal hız (m/s); dz: ortalama düşey yer değiştirme (m); |X|: ortalama mutlak yanal konum (m); |P|: ortalama mutlak toplam eklem gücü (W); |P|/Vy: birim hız başına güç (J/m); $|\tau|$: ortalama mutlak toplam eklem torku (N·m); $|\tau|/Vy$: birim hız başına tork (N·m·s/m). Değerler medyan $\mp$IQR olarak raporlanmıştır; her politika tohumu beş arazi tohumu üzerinde değerlendirilmiştir 91

1. GİRİŞ

Pekiştirmeli öğrenme (Reinforcement Learning, RL), bir ajanın durumları eylemlere eşleyerek sayısal bir ödül sinyalini eniyilemesini öğrendiği bir makine öğrenmesi yaklaşımıdır. Denetimli öğrenmeden farklı olarak ajana hangi eylemi yapması gerektiği açıkça söylenmez; bunun yerine ajan ortamı keşfeder ve daha yüksek ödül üreten eylemleri belirler (Sutton ve Barto 2018). RL algoritmaları dinamik ortamlar içinde çalışır ve en iyi sonuca götüren en uygun eylem dizisini keşfetmeyi amaçlar (Nagabandi vd. 2018). Ajan, ortamla sürekli etkileşim kurarak politikasını öğrenir ve gelecekteki performansını artırmak için bu politikayı giderek iyileştirir (Padakandla 2020).

RL'de ajan politikası, gözlemleri (durum girdileri) eylemlere (çıktılara) eşleyen bir fonksiyondur. Gözlemsel veriler alındığında politika, uygulanacak en uygun eylemi belirler. Yürüyen insansı robotlar için bu gözlemsel veriler tipik olarak eklem açıları, gövde ivmesi ve açısal hızı, ayak temas bilgisi vb. bileşenleri içerir. Politika, bu algısal girdiyi işleyerek motor komutları üretir ve robotun kol ve bacaklarını buna göre hareket ettirir (Lin ve Mitchell 1993; McCallum 1996). Ortam daha sonra seçilen eylemin niteliğini yansıtan sayısal bir ödül döndürür. RL algoritması, gözlemler, gerçekleştirilen eylemler ve alınan ödüllere dayanarak gelecekte daha iyi kararlar verebilmek için politikayı günceller (Zhang vd. 2021). Sonuç olarak ajan, herhangi bir durumda uzun vadede kümülatif ödülü maksimize edecek (eniyileyecek) en iyi eylemi seçmeyi hedefler (Sutton ve Barto 2018; Liu vd. 2020).

RL'de ajan mimarisi içindeki sinir ağları (Neural Network, NN), genellikle aktör-eleştirmen (actor-critic) yapısı kullanılarak hem politikayı hem de değer fonksiyonunu temsil eder (Sutton ve Barto 2018). Aktör ağı, mevcut duruma göre eylem üretir ve böylece ajanın politikasını tanımlar. Eleştirmen ağı ise bu eylemleri, mevcut durum-eylem çifti için Q-değerini kestirerek değerlendirir. Aktör bir eylem seçip ortama uyguladığında, eleştirmen beklenen ödülü (Q-değeri) tahmin ederek bu eylemin kalitesini değerlendirir (Xi vd. 2021). Daha sonra eleştirmen, ortamdan gelen ödülü kullanarak değer kestiriminin doğruluğunu sınar. Oluşan hata, eleştirmenin önceki kestirimi ile anlık ödül ve yeni duruma dayalı güncellenmiş kestirim arasındaki farktır. Eleştirmen bu hatayı kullanarak değer fonksiyonunu günceller ve gelecekteki kestirimlerini iyileştirir. Aktör ise eleştirmenden gelen geri bildirimle parametrelerini güncelleyerek, beklenen kümülatif ödülü artıracak eylemleri seçen politikasını iyileştirir (Grondman vd. 2012). Bu yapı, politikanın yalnızca ham ödüllere dayanmak yerine eleştirmenin sağladığı gradyan bilgisi doğrultusunda eniyilenmesine olanak tanır (Liu vd. 2020).

İnsansı robotik yürüyüş (lokomosyon) için çevre modellemesinde robotun dinamik ve kinematik modeli, eyleyici yapılandırmaları, uzuv uzunlukları ve zemin koşulları gibi fiziksel bileşenler tanımlanır. Buna karşılık modelsiz derin RL (Deep RL, DRL) algoritmaları, sistem dinamiklerini veya çevre geçişlerini açık biçimde modellemeksizin, ortamla etkileşimden elde edilen deneyimler üzerinden denetim politikası öğrenir (Sutton ve Barto 2018). DRL, robotik denetimin çeşitli zorlu alanlarında başarıyla uygulanmış (Gu vd. 2016; Haarnoja vd. 2018a) ve tatmin edici performans göstermiştir (Haarnoja vd. 2019).

Modelsiz sistemlerde ajan tasarımı uygulamaya bağlı olarak önemli ölçüde değişebilir. Bu çalışmada, sürekli eylem uzayları için yaygın olarak kullanılan üç RL algoritması kullanılmış ve karşılaştırılmıştır (Olaz vd. 2023): Derin Deterministik Politika Gradyanı (Deep Deterministic Policy Gradient, DDPG), İkiz Gecikmeli Derin Deterministik Politika Gradyanı (Twin Delayed Deep Deterministic Policy Gradient, TD3) ve Yumuşak Aktör-Eleştirmen (Soft Actor-Critic, SAC). Bu ajanların performans karşılaştırmaları ve değerlendirmeleri sonraki bölümlerde sunulmaktadır.

DDPG ve TD3 algoritmaları, sürekli eylem uzaylarına sahip ortamlar için tasarlanmış, davranıştan bağımsız (off-policy) aktör-eleştirmen yöntemleridir. Burada davranıştan bağımsızlık, ajanın politika güncellemelerini yalnızca güncel politika altında elde edilen verilerle değil, daha önce toplanmış ve yeniden oynatma arabelleğinde saklanan deneyimler üzerinden de gerçekleştirebilmesini ifade eder. DDPG, Q-değerini kestirmek için bir eleştirmen ağı kullanır; bu sayede deterministik bir politika öğrenir. Eğitim kararlılığını artırmak amacıyla hedef aktör ve hedef eleştirmen ağları ile deneyim yeniden oynatma arabelleğinden yararlanır (Fujimoto vd. 2018; Lillicrap vd. 2015). TD3, DDPG ile benzer bir yapıyı paylaşmakla birlikte hedef ağlar ve yeniden oynatma arabelleğini yine kullanır; ancak iki eleştirmen ağı kullanarak Q-değeri kestirimini daha sağlam hale getirir ve böylece kararlılığı ile başarımı iyileştirir (Fujimoto vd. 2018).

SAC algoritması ise hem sürekli hem de ayrık eylem uzayları için tasarlanmış, davranıştan bağımsız (off-policy) bir aktör-eleştirmen yöntemidir. Off-policy yapı sayesinde SAC, güncel politikadan bağımsız olarak önceden toplanmış deneyimler üzerinden öğrenebilir. SAC, beklenen kümülatif ödül ile politika entropisi arasındaki ödünleşimi gözeten stokastik bir politika öğrenir. Entropi terimi, belirli bir durumda politikanın eylem dağılımındaki belirsizliği ölçer; daha yüksek entropi daha fazla keşfi teşvik eder ve eğitim sürecinde sağlamlığı artırır (Haarnoja vd. 2018b; Sutton ve Barto 2018). SAC, politika değerlendirmesinin doğruluğunu artırmak için iki eleştirmen ağı kullanır; ayrıca hedef eleştirmenler ve deneyim yeniden oynatma arabelleğinden yararlanır (Fujimoto vd. 2018).

Zaman içinde bu temel algoritmalar, güçlü yönlerinden yararlanmak ve sınırlılıklarını azaltmak amacıyla tamamlayıcı stratejiler ve algoritmik yapılarla birleştirilmiştir. Özellikle, bu ajanları evrimsel stratejilerle birleştiren hibrit yaklaşımlar önerilmiş ve çeşitli ortamlarda performansı artırdığı gösterilmiştir (Khadka ve Tumer 2018; Pourchot ve Sigaud 2019).

Evrimsel algoritmalar (EA), doğadaki doğal seçim ve genetik işlemlerden esinlenen bir arama algoritmaları sınıfıdır. Bu algoritmalar üç temel işleme dayanır: yeni çözümler üretme, mevcut çözümleri değiştirme (mutasyon ve çaprazlama) ve en iyi çözümleri seçme. Bu işlemler aday çözümlerden oluşan bir popülasyona yinelemeli olarak uygulanır. Süreç boyunca başarımı yüksek çözümler korunurken yeni adaylar sürekli olarak popülasyona dahil edilir (Khadka ve Tumer 2018). Seçim mekanizması çoğunlukla olasılıksaldır ve çözüm kalitesinin göstergesi olan daha yüksek uygunluk (fitness) değerlerine sahip çözümleri tercih eder. Jenerasyonlar ilerledikçe popülasyonun genel kalitesinin artması beklenir (Fogel 2006; Spears vd. 1993).

Evrimsel Strateji (ES) ile RL'nin birleştirilmesiyle geliştirilen ajan (ES-RL), hibrit bir eğitim yapısı kullanır. ES-RL, çekirdek bileşeni tekrar oynatma arabelleği sistemi olan

Evolutionary RL (ERL) çerçevesinden türemektedir. Pourchot ve Sigaud (2019) tarafından önerilen bir örnekte, Cross-Entropy Method (CEM) politika tabanlı olmayan RL teknikleriyle birleştirilmiştir. ERL'de aktör ağlarından oluşan bir popülasyon standart EA yapıları ile eniyilenir ve ortaya çıkan deneyimler tekrar oynatma arabelleği içinde depolanır. Bu paylaşılan deneyim verisi hem aktör hem de eleştirmen ağlarını eğitmek için kullanılarak farklı davranışsal politikalardan RL modeline bilgi aktarımı sağlanır (Pourchot ve Sigaud 2019; Callaghan vd. 2023).

ES-RL algoritmalarının SAC, DDPG ve TD3 gibi aktör-eleştirmen tabanlı yöntemlerle birleştirildiği yapılarda, başlangıçta rastgele başlatılmış parametrelere sahip bir aktör popülasyonu oluşturulur ve buna ek olarak gradyan tabanlı öğrenme gerçekleştiren ayrı bir aktör-eleştirmen bileşeni tanımlanır (Bai vd. 2023; Abdeen vd. 2024). Aktör popülasyonundaki bireyler ortamla etkileşime girerek eylemler üretir, bu etkileşimlerden elde edilen durum, eylem, ödül ve bir sonraki durum geçişleri yeniden oynatma arabelleğinde saklanır ve söz konusu deneyimler daha sonra SAC, DDPG veya TD3 güncellemeleri aracılığıyla aktör ve eleştirmen ağlarının iyileştirilmesinde kullanılır (Lee vd. 2021; Roshan vd. 2023). Her bir aktör bireyinin uygunluğu, genellikle bir bölüm boyunca elde edilen toplam ödüle göre belirlenir ve bu değer bireyin başarım ölçütü olarak kullanılır. Seçim aşamasında bireyler uygunluk puanlarına göre sıralanır, en iyi başarımı gösteren bireyler elit olarak korunur ve bu elitlerin parametrelerinde küçük değişiklikler yapılarak yeni bir aktör popülasyonu oluşturulur. Böylece evrimsel arama, popülasyon düzeyinde geniş bir çözüm uzayını tararken, SAC, DDPG veya TD3 tabanlı gradyan güncellemeleri seçilen politikaları daha ince ölçekte iyileştirir. Değerlendirme, seçim, yeni bireyler üretme ve gradyan tabanlı güncelleme adımlarından oluşan bu döngü, yakınsama sağlanıncaya ya da önceden tanımlanmış durdurma ölçütüne ulaşılıncaya kadar yinelemeli olarak sürdürülür (Müller ve Glasmachers 2018; Bai vd. 2023).

Bu çalışmada değerlendirme, ödül fonksiyonunun sayısal değerinin ötesine taşınarak lokomosyon güvenilirliğini ve yürüyüş dinamiklerinin yapısal özelliklerini doğrudan etkileyen performans ölçütlerini de kapsamaktadır. Özellikle, ödül fonksiyonunun bileşenleri arasında yer alan tork verimliliği, yanal ve düşey sapmalar ile doğrusal hız gibi büyüklükler ve Q-değeri kestirimleri tamamlayıcı göstergeler olarak incelenmiştir. Buna ek olarak, ikinci aşama analizlerde eklem davranışının kararlı durumdaki periyodik yapısını değerlendirmek amacıyla olay hizalamalı limit-çevrim faz portreleri (yoğunluk konturları) kullanılmıştır. Bu kapsamda, eklem açısı ile eklem açısal hızı arasındaki ilişki, kolların salınıma katıldığı iki farklı kol koordinasyon düzeninde ve kolların sabit tutulduğu durumda karşılaştırılmıştır; alt ekstremite eklemleri (kalça, diz, ayak bileği) için ise tüm kol konfigürasyonları birlikte değerlendirilmiştir. Böylece yalnızca sayısal ödül ve özet metrikler değil, aynı zamanda yürüyüşün çevrimsel düzenliliği, adımlar arası değişkenlik ve eklem koordinasyon örüntülerindeki farklılıklar da değerlendirmeye dahil edilmiştir. Bu parametreler ve faz-portresi temelli analizler birlikte ele alındığında, insansı robot lokomosyonunda denetleyici ajanın performansına daha geniş ve çok boyutlu bir bakış açısı sunulmaktadır.

Bu tez çalışmasında izlenen yöntem iki aşamalı bir çerçevede yapılandırılmıştır. İlk aşamada, ES-RL ailesi içinde yer alan ES-SAC, ES-TD3 ve ES-DDPG ajanları aynı problem kurulumu altında karşılaştırılmış; öğrenme performansı, kararlılık ve yürüyüşe ilişkin temel çıktı metrikleri üzerinden en uygun ajan belirlenmiştir. Bu karşılaştırmalar sonucunda elde edilen bulgular doğrultusunda, izleyen aşamada en iyi performansı veren

ajan temel denetleyici olarak seçilmiş ve daha ayrıntılı biyomekanik/denetim odaklı analizler bu ajan üzerinden yürütülmüştür. Bu kapsamda, seçilen ajan kullanılarak kol salınım stratejilerinin insansı robot yürüyüşü üzerindeki etkileri sistematik biçimde incelenmiştir.

İnsanlar yürürken tipik olarak kollarını sallama eğilimindedir. Kol salınımı kas etkinliği gerektirir ve enerji açısından maliyetli olabilir. Vücudun diğer bölümlerinde ortaya çıkabilecek olası yararların bu maliyeti telafi edip etmediği ise açık değildir. İnsansı robotların tasarımları da benzer biçimde insan yürüyüşünden esinlenmektedir (Collins vd. 2009). Ancak insansı yürüyüş denetimine yönelik çoğu yaklaşım (Singh vd. 2023; Wu vd. 2023; Itahashi vd. 2024; Li vd. 2024), bacak hareketine odaklanmış; üst gövde ve kol eklemlerinin katkısını arka planda bırakmıştır. Bu durum önemlidir; çünkü kol koordinasyonu, tüm vücut dengesinin korunmasını, kinematik izleme doğruluğunu ve enerjiyle ilişkili çabanın dağılımını aynı anda etkileyebilmektedir. Bu unsurların tümü, insansı robotlarda kararlı, verimli ve uygulanabilir bir yürüyüşün temel bileşenlerini oluşturmaktadır.

İnsansı robot yürüyüşünün incelenmesinde kütle merkezi (Center of Mass, CoM) ve Doğrusal Ters Sarkaç Modeli (Linear Inverted Pendulum Model, LIPM) gibi kavramlar, yürüyüş dinamiğini anlamada temel bir çerçeve sunar. CoM, robotun toplam kütlesinin eşdeğer olarak toplandığı noktayı ifade eder ve yürüyüş sırasında bu noktanın ileri, yanal ve düşey yöndeki hareketi denge, salınım ve kararlılık hakkında önemli bilgi sağlar. LIPM ise yürüyüşü sadeleştirilmiş bir ters sarkaç yaklaşımıyla açıklayan kavramsal bir modeldir; özellikle denge davranışının ve gövde hareketinin temel eğilimlerini yorumlamada kullanılır (Li vd. 2021). Bununla birlikte, bu çalışmada denetim doğrudan LIPM üzerine kurulmamış, daha yüksek fiziksel gerçekçiliğe sahip çok-cisimli bir simülasyon modeli tercih edilmiştir. Bu çalışmada kullanılan insansı robot modeli, Simulink/Simscape Multibody ortamında oluşturulmuş torkla sürülen çok cisimli bir sistemdir. Modelde alt ekstremite, üst ekstremite ve gövde bileşenleri birlikte temsil edilmekte; eklem hareketleri, gövde duruşu ve ayak-zemin etkileşimi eşzamanlı olarak simüle edilmektedir. Bu sayede yürüyüş yalnızca ödül eğrileri üzerinden değil, aynı zamanda hız, sapma, tork, güç ve dengeyle ilişkili fiziksel göstergeler üzerinden de değerlendirilebilmektedir. RL ajanı, çevreden aldığı gözlemler (eklem bilgileri, gövde durumu, temas bilgisi vb.) üzerinden eklem tork komutları üretmekte; fiziksel model bu komutlara karşılık robotun hareketini oluşturmaktadır. Elde edilen veriler hem ödül hesabında hem de yürüyüş performans analizinde kullanılmaktadır. Böylece çalışma, bir yandan CoM ve LIPM gibi temel yürüyüş kavramlarıyla ilişki kurulmasına imkân verirken, diğer yandan ayrıntılı fiziksel model üzerinde veri temelli bir denetim ve değerlendirme altyapısı sunmaktadır.

Bu çalışmanın amacı, DRL tabanlı bir denetleyici kullanarak kol salınımının insansı robotların yürüyüş performansını nasıl etkilediğini ayrıntılı biçimde incelemektir. Bu amaçla üç kol-model senaryosu karşılaştırılmıştır. Birinci senaryoda kollar tamamen sabitlenmiş ve kol salınımına izin verilmemiştir. İkinci senaryoda kollar, aynı taraftaki bacaklarla aynı fazda salınmıştır (örneğin sağ bacak ileri giderken sağ kolun da ileri gitmesi). Üçüncü senaryoda ise kollar, doğal insan yürüyüşündeki gibi karşı taraftaki bacaklarla eşzamanlı ters fazda salınmıştır (örneğin sağ bacak öne adım atarken sol kolun öne salınması). Bu üç koşul boyunca yürüyüş kararlılığı, enerji harcaması ve dengeyi belirleyen çeşitli ölçütler analiz edilmiştir. Böylece kol hareketinin iki ayaklı yürüyüş

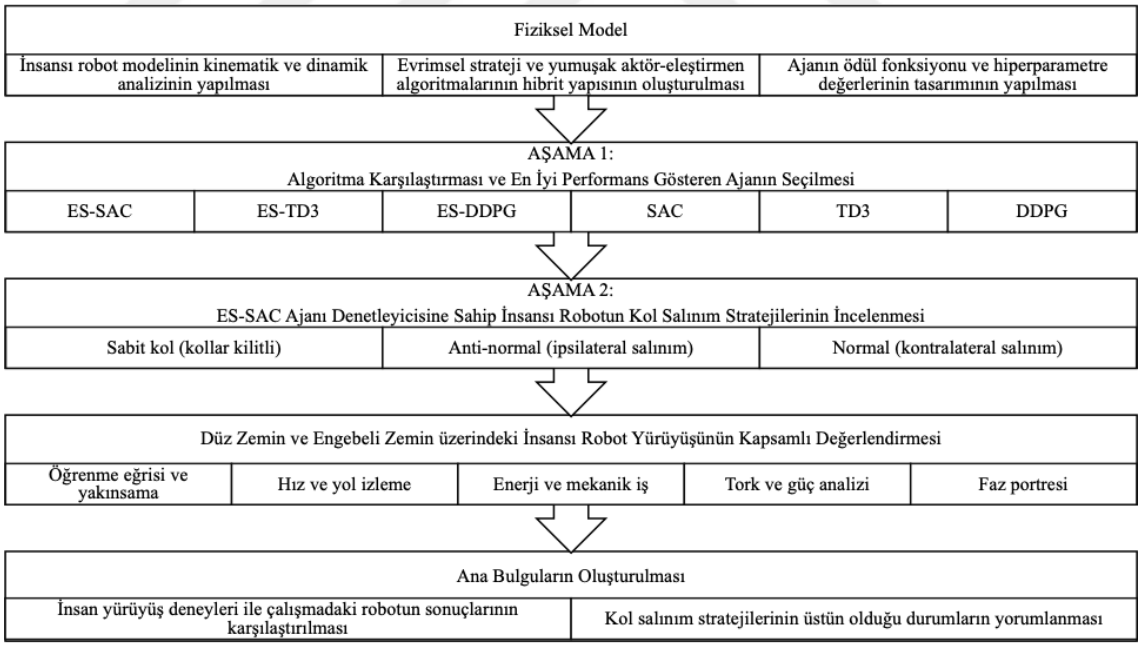
dinamikleri üzerindeki etkileri sayısal kanıtlarla nicel olarak ortaya konmuştur. Ayrıca bu karşılaştırma, kol hareketini yalnızca estetik bir serbestlik derecesi (Degrees of Freedom, DoF) olarak ele almak yerine, her bir kol stratejisinin kararlılık-hız davranışı ile çaba/ekonomi açısından göreli üstünlük ve sınırlılıklarını görünür kılacak şekilde tasarlanmıştır. Burada DoF, bir mekanik sistemin uzayda bağımsız olarak hareket edebileceği yönlerin sayısını ifade eder; örneğin bir kolun öne-arkaya, yana ve dönme hareketleri farklı serbestlik derecelerine karşılık gelir. Ters fazlı örüntü tipik insan koordinasyonuna benzerken aynı fazlı örüntü buna benzemediği için, analiz aynı zamanda insansı denetim davranışının insan-benzeri yürüyüş koordinasyonuyla hangi noktalarda örtüştüğünü veya ayrıştığını da vurgulamaktadır.

Collins vd. (2009) çalışmasına göre yürüyüş sırasında kol salınımı önemli işlevler görmektedir. Bu çalışma, kolların sabitlenmesinin yürüyüşün metabolik enerji maliyetini yaklaşık %12 artırdığını göstermiştir. Kollar bacaklarla aynı tarafta aynı fazda salındığında enerji kullanımı %26'ya kadar artabilmektedir. Bir diğer dikkat çekici bulgu, kol salınımının ayak-zemin teması sırasında düşey eksen etrafındaki zemin tepki momentini belirgin biçimde azaltmasıdır. Kollar tamamen sabitlendiğinde bu moment, normal kol salınımına göre yaklaşık %63 artmaktadır. Normal kol salınımı için gereken omuz torkunun da oldukça küçük olduğu gösterilmiştir; bu da düşük bir kas maliyetine karşılık kolların hem dengeye hem de yürüyüşün enerji verimliliğine katkı sağladığını göstermektedir. Bununla birlikte, bu bulgular insansı robotlara doğrudan genellenemeyebilir; çünkü robotun yapısal özellikleri, eyleyici sistemi ve denetleyici hedefleri, insan benzeri koordinasyon ile göreve en uygun koordinasyon arasındaki ödünleşimi değiştirebilir.

DRL yöntemleri insansı yürüyüşte tatmin edici sonuçlar üretmiştir (Heess vd. 2017; Fujimoto vd. 2018; Lillicrap vd. 2015). Güncel çalışmalar, kolların kullanılmasının yürüyüş hızı ve dengeyi belirgin biçimde iyileştirebildiğini göstermektedir. Özellikle daha yüksek yürüyüş hızlarında bacak salınımı dengesiz açısal momentum üretebilir ve robotu tökezlemeye yatkın hale getirebilir (Zhang vd. 2024; Wang vd. 2025). Bu soruna çözüm olarak Zhang vd. (2024), DRL tabanlı bir denetleyicinin ödül fonksiyonuna açısal momentum minimizasyon terimi eklemiş; böylece robotun düşey eksen (yaw) etrafındaki açısal salınımını dengelemek ve yürüyüşü kararlı hale getirmek için kollarını aktif biçimde kullanmasını sağlamıştır.

Önceki çalışmalar (Collins vd. 2009; Herr ve Popovic 2008; Umberger 2008), kol salınımının düşey düzlemde CoM hareketini azaltabildiğini, tüm-vücut açısal momentumu dengeleyebildiğini ve düşey eksen etrafındaki zemin tepki momentini azaltabildiğini göstermektedir. Kollar bacaklarla ters fazda hareket ettiğinde, bacakların ürettiği açısal momentuma karşı etki oluşturarak gövdedeki kontrolsüz dönmeleri ve aşırı yanal salınımı azaltırlar. Bu içgörüler, yalnızca ödül eğrilerine dayanmak yerine hem kinematik performansı hem de çaba/enerji kullanımını yansıtan metriklerle, insansı robotta kol koordinasyonu stratejilerinin kontrollü bir karşılaştırmasını motive etmektedir. Bu çerçevede, sabit bir denetleyici ve ortak bir eğitim düzeni altında farklı kol koordinasyonu stratejilerinin kararlı durum yürüyüş kalitesi, enerjiyle ilişkili maliyet vekil ölçütleri ve sağlamlıkla ilişkili denge göstergeleri üzerindeki etkilerinin belirlenmesi temel araştırma sorusunu oluşturmaktadır. Buna ek olarak, bu performans ölçütleri arasındaki ödünleşimin hangi strateji tarafından daha dengeli biçimde yönetildiğinin ortaya konması amaçlanmaktadır.

Çalışmanın ikinci aşamasında, birinci aşamada en başarılı sonucu üreten ES-SAC algoritması kullanılarak tek bir ajan mimarisi benimsenmiştir. ES ve SAC bileşenleri sırasıyla (Khadka ve Tumer 2018; Pourchot ve Sigaud 2019) ve (Heess vd. 2017; Fujimoto vd. 2018; Lillicrap vd. 2015) çalışmalarından esinlenerek geliştirilmiştir. Bu ajan ile hem bacak hem de kol hareketleri denetlenmekte ve ödül fonksiyonu buna uygun biçimde, ileri yönlü ilerlemeyi ve düşmeden kaçınmayı hedeflerken aynı zamanda kararlılık ve verimlilikle ilişkili davranışları açıkça teşvik eden çok bileşenli bir amaç fonksiyonu olarak tasarlanmaktadır. Öğrenme eğrilerinin ötesinde, kol koordinasyonunun yürüyüş dinamiklerini ölçülebilir şekilde nasıl değiştirdiğini yorumlamak için ayrıntılı bir kararlı durum yürüyüş analizi yapılmıştır. Buna paralel olarak, kol salınım modellemesinin öğrenme sürecinin kendisi üzerindeki etkisi de incelenmiş; her kol stratejisinin ES-SAC altında eğitim davranışını nasıl şekillendirdiği, yakınsama eğilimleri, tohumlar arası değişkenlik ve ödül/Q-değeri gibi değere ilişkin öğrenme sinyalleri üzerinden değerlendirilmiştir. Ayrıca koşul başına beş bağımsız çalıştırmadan oluşan çoklu-tohum tasarımı kullanılarak çalıştırmaya özgü dalgalanmaların azaltılması ve kararlılık ile genellenebilirliğin daha güvenilir biçimde değerlendirilmesi amaçlanmıştır. Sonuç olarak, biyomekanik bulguları ES-SAC denetimi ile bir araya getirerek ve sabit, aynı fazlı ve ters fazlı kol stratejilerini aynı eğitim düzeni altında karşılaştırarak, kol hareketinin insansı lokomosyonda yürüyüş kararlılığı, denge ve enerjiyle ilişkili performans arasındaki ödünleşimleri nasıl şekillendirdiğine ilişkin yapılandırılmış bir değerlendirme sunulmuştur.



Şekil 1.1. Çalışmada izlenen iki aşamalı deneysel yöntemin genel akış şeması

Şekil 1.1, bu tez çalışmasında baştan sona izlenen deneysel süreci iki ana aşama üzerinden özetlemektedir. Süreç, insansı robotun fiziksel modelinin kurulmasıyla

başlamaktadır; bu kapsamda robotun kinematik ve dinamik yapısı tanımlanmış, evrimsel strateji ES ile yumuşak aktör-eleştirmen SAC algoritmalarının birleştirildiği hibrit öğrenme çerçevesi oluşturulmuş, ödül fonksiyonu ve hiperparametre değerleri tasarlanmış ve tüm bu kurgu, Matlab ortamında ES ile eğitimi destekleyen ajan ailesi (SAC, TD3, DDPG ve bunların ES tabanlı hibrit sürümleri) esas alınarak yapılandırılmıştır. Birinci aşamada altı farklı ajan (ES-SAC, ES-TD3, ES-DDPG, SAC, TD3, DDPG) aynı koşullar altında eğitilerek öğrenme başarımı, kararlılık ve temel yürüyüş ölçütleri üzerinden karşılaştırılmış ve en tutarlı performansı gösteren ES-SAC, ikinci aşamanın denetleyicisi olarak belirlenmiştir. İkinci aşamada odak noktası kol hareketine yönelmektedir: kolların tamamen kilitlendiği sabit kol durumu, kolların aynı taraftaki bacakla birlikte salındığı anti-normal (ipsilateral) durum ve insanların doğal yürüyüşündeki gibi kolların karşı taraftaki bacakla eşzamanlı hareket ettiği normal (kontralateral) durum olmak üzere üç farklı kol salınım stratejisi hem düz hem de engebeli zeminde ayrı ayrı eğitilmiştir. Eğitimin ardından tüm stratejiler öğrenme eğrisi yakınsaması, ileri hız, yanal ve düşey sapma, eklem torku ve gücü, mekanik iş ile faz portresi gibi çok yönlü ölçütler üzerinden kapsamlı biçimde değerlendirilmiştir. Son olarak elde edilen robotik bulgular, insan yürüyüş deneylerinden elde edilen sonuçlarla karşılaştırılarak kol salınım stratejilerinin üstün olduğu koşullar yorumlanmıştır. Bu iki aşamalı sürecin nihai amacı, kol salınımının insansı robot yürüyüşünde yalnızca görsel bir ayrıntı olmadığını ortaya koyarak öğrenme verimliliğinden enerji kullanımına, kararlılıktan hız başarımına kadar geniş bir alanda belirleyici bir etken olduğunu sistematik deneysel kanıtlarla göstermektir.

2. KAYNAK TARAMASI

İnsansı robot lokomosyonu, doğrusal olmayan çok cisimli yapısı, süreksiz ayak-yer teması geçişleri, yüksek boyutlu durum-eylem uzayları ve denge ile enerji verimliliği arasındaki karşıt gereksinimler nedeniyle robotik denetimin en zorlu problemlerinden biri olarak kabul edilmektedir. Bu nedenle, alanda bir yandan azaltılmış dereceli dinamik modeller, ZMP, LIPM ve açısal momentum tabanlı denetim yaklaşımları geliştirilmiş, diğer yandan özellikle son yıllarda DRL tabanlı yöntemler doğrudan çok cisimli sistemler üzerinde etkili bir alternatif haline gelmiştir (Vukobratović ve Borovac 2004; Kajita vd. 2001; Mitobe vd. 2004; Kotha vd. 2024). Ayrıca güncel derleme çalışmaları (Tang vd. 2024; Bao vd. 2026), iki ayaklı robot lokomosyonu alanındaki DRL tabanlı yöntemlerin artık yalnızca düz zemin yürüyüşü ile sınırlı kalmadığını; hız takibi, arazi koşullarına genelleme, hiyerarşik denetim, simülasyondan gerçek sisteme aktarım ve birden fazla hareket yeteneğini kapsayan lokomosyon görevleri gibi daha karmaşık problemlere yöneldiğini göstermektedir. Bao vd. (2026), DRL tabanlı iki ayaklı robot lokomosyon literatürünü uçtan uca ve hiyerarşik yaklaşımlar olarak sınıflandırmış, her iki çerçevenin güçlü ve zayıf yönlerini sistematik biçimde tartışmıştır. Benzer şekilde Tang vd. (2024), gerçek dünya robot uygulamaları açısından DRL'nin özellikle örnek verimliliği, kararlılık ve değerlendirme protokolleri bakımından hâlen önemli güçlükler içerdiğini, buna karşın karmaşık robot davranışlarının öğrenilmesinde güçlü bir araç niteliği taşıdığını vurgulamıştır.

İnsansı yürüyüşte kol salınımının rolü, yalnızca biyomekanik açıdan değil, robot denetimi açısından da önemli bir araştırma konusu olarak ele alınmaktadır (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010). İnsan yürüyüşünde kol hareketleri uzun süre ikincil bir unsur olarak değerlendirilmiş olmakla birlikte, deneysel biyomekanik çalışmalar bu hareketlerin işlevsel bir katkı taşıdığını ortaya koymuştur. Pontzer vd. (2009), kol salınımını pasif bir kütle sönümleme mekanizmasına benzer biçimde incelemiş ve kolların gövde ile baş hareketlerindeki dönmeyi azaltıcı bir işleve sahip olduğunu göstermiştir. Söz konusu çalışma, kol salınımının alt gövde hareketleriyle güçlü biçimde bağlantılı olduğunu ve üst gövde rotasyonlarını sınırlandırdığını da belirtmektedir. Herr ve Popovic (2008) ise tüm vücut açısal momentumunun insan yürüyüşü boyunca üç eksende sıkı biçimde düzenlendiğini göstermiş, yürüyüş kararlılığının yalnızca ayak yerleşimiyle değil, segmentler arasındaki momentum dengesiyle de ilişkili olduğunu ortaya koymuştur. Bruijn vd. (2010) çalışmasında kol salınımının özellikle dış bozucu etkiler sonrasındaki toparlanma davranışında önemli olduğunu, kararlı rejimde sürdürülen yürüyüş sırasında yerel kararlılık üzerindeki etkisinin daha sınırlı kalabilmesine karşın, bozucu etki sonrası normal yürüyüş düzenine dönüş hızını etkilediğini bildirmiştir. Bu çalışmalar birlikte değerlendirildiğinde, kol salınımının temel etkisinin yalnızca enerji maliyetini azaltmak olmadığı; aynı zamanda gövde rotasyonlarını sınırlama, açısal momentumu dengeleme ve bozucu etkiler sonrasındaki toparlanmaya katkı sağlama olduğu anlaşılmaktadır (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010). Bu çerçevede, sabit kol, aynı fazlı kol ve ters fazlı kol stratejilerinin karşılaştırılması, literatürde yer alan biyomekanik bulguların robotik sistemler bağlamında incelenmesine doğrudan karşılık gelmektedir.

İnsan yürüyüşünün sagittal ve frontal düzlemdeki temel kinematik özellikleri, adım çevrimi boyunca uzuv segmentlerinin ardışık konumları üzerinden açıklanabilmektedir (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010). Bu

kapsamda, Şekil 2.1'de aktif denetim bakımından sınırlı sayıda serbestlik derecesiyle modellenmiş insansı robotun yürüyüşü yandan görünüşte sunulmakta; alt ekstremitenin ileri geri salınımı, eklemli bacak zincirinin çevrim boyunca aldığı konumlar ve ayak-yer teması ile ilişkili ilerleme düzeni izlenebilmektedir.



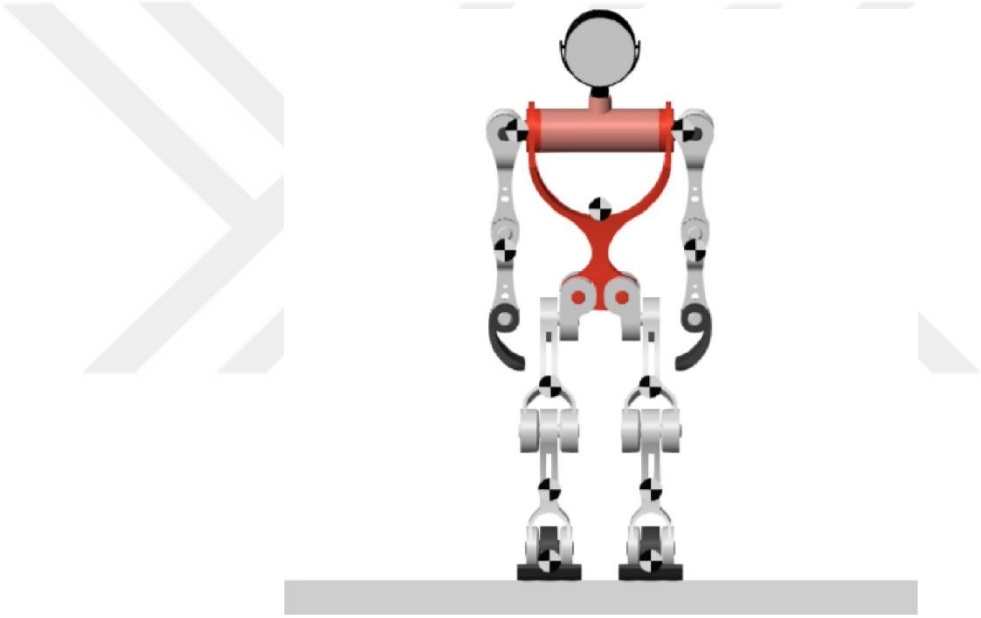
Şekil 2.1. Çalışmada kullanılan insansı robot modelinin yandan görünüşü

Şekil 2.2'de ise robotun önden görünüşü verilmekte; gövdenin orta hat etrafındaki duruşu, alt ekstremitelerin yanal yerleşimi ve üst ekstremitelerin gövdeye göre konumu açık biçimde görülmektedir.

Şekil 2.3'te ise yürüyüş hem yandan hem de önden görünüşte ardışık uzuv konumlarıyla şematik olarak gösterilmekte; böylece hareketin çevrimsel yapısı ile sagittal (Şekil 2.3a) ve frontal (Şekil 2.3b) düzlem koordinasyonu birlikte ortaya konmaktadır (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010). Özellikle Şekil 2.3'te yer alan kesikli konturlar, uzuvların yürüyüş çevrimi boyunca aldığı ardışık konumları temsil etmekte ve insan yürüyüşünün periyodik karakterini görsel olarak açıklamaktadır. Bu bakımdan söz konusu şekiller, yalnızca yürüyen bir sistemin dış görünüşünü sunan görseller olarak değil, yürüyüş hareketinin iki temel düzlemdeki kinematik bileşenlerini açıklayan tamamlayıcı görsel temsiller olarak ele alınmalıdır.

Buna karşılık Şekil 2.1 ve Şekil 2.2, yalnızca yürüyüşün görsel sunumu açısından değil, aynı zamanda çalışmada kullanılan insansı robot modelinin yapısal özelliklerini göstermesi bakımından da önem taşımaktadır. Güncel bipedal lokomosyon literatüründe, özellikle DRL tabanlı yaklaşımlarda, denetim probleminin yönetilebilir hale getirilebilmesi amacıyla robot modellerinin aktif denetim bakımından sadeleştirilmiş DoF yapılarıyla ele alındığı görülmektedir. Bu çalışmada kullanılan model de 3 boyutlu (3B) uzayda çok cisimli bir sistem olarak tanımlanmış olmakla birlikte, aktif denetim

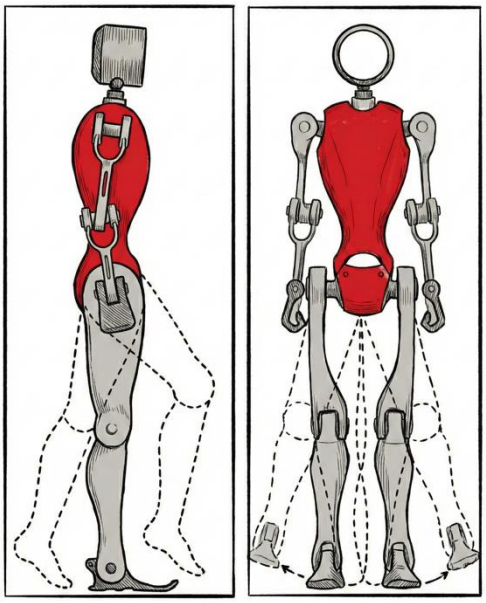
bakımından sınırlı sayıda DoF içerecek biçimde düzenlenmiştir. Bu kapsamda alt ekstremitelerde her bir bacak için üçer aktif dönel eklem tanımlanmış, üst ekstremitelerde ise kol salınımı yalnızca omuz eklemleri üzerinden modellenmiştir. Böylece, insan biyomekaniğinden esinlenen temel yürüyüş özellikleri korunurken, öğrenme ve denetim sürecini daha izlenebilir ve hesaplama açısından daha uygulanabilir kılan sadeleştirilmiş bir insansı robot modeli elde edilmiştir (Radosavovic vd. 2023; Li vd. 2024; Bao vd. 2026). Sonuç olarak, Şekil 2.1 ve Şekil 2.2 robot modelinin yapısal özelliklerini ve DoF kurgusunu, Şekil 2.3 ise insan yürüyüşünün biyomekanik ve kinematik temelini yansıtan bir görsel bütünlük sunmaktadır. Bu nedenle üç şekil birlikte değerlendirildiğinde, hem insan yürüyüşüne ilişkin biyomekanik arka plan hem de bu arka planın sadeleştirilmiş bir insansı robot modeli üzerinde nasıl temsil edildiği literatürle uyumlu biçimde açıklanabilmektedir (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010; Pourchot ve Sigaud 2019; Radosavovic vd. 2023; Li vd. 2024).



Şekil 2.2. Çalışmada kullanılan indirgenmiş serbestlik dereceli insansı robot modelinin önden görünüşü

RL alanında, sürekli eylem uzaylarına sahip kontrol problemlerinde yaygın olarak başvurulan yöntemlerden biri olan DDPG, aktör-eleştirmen yapısını derin sinir ağı (Deep NN, DNN) tabanlı fonksiyon yaklaşımı ile birleştiren temel yaklaşımlardan biri olmuştur. Bununla birlikte, sonraki çalışmalar DDPG'nin değer fonksiyonunda ortaya çıkan aşırı değer tahmini ve eğitim sürecindeki kararsızlıklar nedeniyle ayar duyarlılığı yüksek bir yöntem olduğunu göstermiştir (Fujimoto vd. 2018; Lillicrap vd. 2015). Fujimoto vd. (2018) tarafından geliştirilen TD3 algoritması, bu sorunu çift eleştirmen yapısı ve gecikmeli politika güncellemesi aracılığıyla azaltmayı önermiş; sürekli kontrol problemlerinde aşırı değer tahmini yanlılığının aktör-eleştirmen tabanlı yöntemlerde önemli bir sorun oluşturduğunu ortaya koymuştur. Haarnoja vd. (2018a) tarafından

sunulan SAC yaklaşımı ise maksimum entropi çerçevesi içinde ödül ile keşif arasındaki dengeyi daha sistematik biçimde kurarak daha kararlı bir öğrenme yapısı sağlamıştır. Ayrıca yine çalışmada, yöntemin hem yüksek başarım hem de farklı tohumlar altında daha tutarlı öğrenme davranışı sergilediği gösterilmiştir. Bu nedenle, DDPG, TD3 ve SAC algoritmalarının birlikte ele alınması yalnızca hangi ajanın daha yüksek ödül ürettiğinin belirlenmesi açısından değil, aynı zamanda sürekli yürüme kontrolü problemlerinde hangi yaklaşımın daha kararlı, daha yumuşak ve daha genellenebilir bir denetim davranışı oluşturduğunun incelenmesi açısından da önem taşımaktadır (Fujimoto vd. 2018; Haarnoja vd. 2018a; Lillicrap vd. 2015).



(a) (b)

Şekil 2.3. a) İnsan yürüyüşünün sagittal düzlemde ardışık uzuv konumları; **b)** İnsan yürüyüşünün frontal düzlemde ardışık uzuv konumları

Bu çalışmanın özgün yönü, kol salınımının yürüyüş öğrenmesi üzerindeki etkisini ödül fonksiyonu, öğrenme davranışı, kararlı durum performansı ve insan yürüyüşüyle karşılaştırmalı bulgular düzeyinde bütüncül olarak incelemesidir. Tez kapsamında, baskın olarak sagittal düzlem hareketine sahip indirgenmiş DoF uzaysal bir insansı yürüteç üzerinde normal, anti-normal ve sabit kol olmak üzere üç farklı kol stratejisi hem düz hem de engebeli zeminde karşılaştırılmıştır; her koşul için çoklu rastgele tohum kullanılarak öğrenme eğrileri, yakınsama özellikleri ve eğitim sonrası yürüme ölçütleri sistematik biçimde değerlendirilmiştir. Çalışmada kullanılan ödül yapısı, ortak yürüme hedeflerini korurken kol-bacak koordinasyonuna ilişkin terimlerin koşula bağlı olarak yeniden düzenlenmesi esasına dayanmaktadır. Normal ve anti-normal kol salınımı için hedeflenen koordinasyon biçimi ödül yapısına açık biçimde yerleştirilmiş, sabit kol durumunda ise kola ilişkin şekillendirme terimleri devre dışı bırakılmıştır. Bu sayede, kol salınımı stratejisinin yalnızca son yürüme performansını değil, öğrenme sürecinin

yapısını da nasıl etkilediği doğrudan incelenebilir hale gelmiştir. Ayrıca koşullar arası değerlendirme yalnızca toplam ödül üzerinden yapılmamış; ödülden bağımsız olarak hesaplanan ileri hız, yanal sapma, düşey salınım, güç, tork, adım başına iş, mesafe başına iş ve taşıma maliyeti gibi kararlı durum ölçütleri de birlikte raporlanmıştır. Bu yaklaşım, hangi ajanın veya hangi kol stratejisinin daha yüksek ödül ürettiğinden öte, kol salınımının öğrenme verimliliğini, hareket ekonomisini ve yürüyüş kararlılığını nasıl biçimlendirdiğini daha ayrıntılı biçimde ortaya koymaktadır. Nitekim tez bulguları, normal kol salınımı koşulunun yüksek ödül düzeyine daha erken ulaştığını ve daha kararlı bir yakınsama eğilimi sergilediğini; anti-normal koşulunun ise daha zayıf bir öğrenme profiline sahip olduğunu göstermektedir. Bunun yanında, Collins vd. (2009) insan yürüyüşü deneyleriyle kurulan karşılaştırma üzerinden, doğal kol salınımının daha ekonomik bir rejime yakın davrandığı, anti-normal ve sabit kol koşullarının ise bundan belirgin biçimde ayrıştığı gösterilmiştir. Dolayısıyla bu çalışmanın asıl katkısı, kol salınımı stratejilerinin insansı robotlarda yalnızca yürüyüş çıktısını değil, öğrenme sürecini, enerjiyle ilişkili vekil ölçütleri ve insan biyomekaniğiyle benzerlik ve farklılıklarını birlikte ele alan ayrıntılı bir karşılaştırma çerçevesi sunmasıdır.

Güncel insansı ve iki ayaklı robot yürüyüşü literatürü, DRL tabanlı denetleyicilerin artık yalnızca simülasyon ortamındaki başarılarla sınırlı kalmadığını, gerçek sistemler üzerinde de uygulanabildiğini göstermektedir (Radosavovic vd. 2023; Tang vd. 2024; Li vd. 2024). Radosavovic vd. (2023), gerçek boyutlu bir insansı robot üzerinde tamamen öğrenme tabanlı bir yürüyüş denetleyicisi sunmuş; bu denetleyicinin dış bozucular karşısında dayanıklı olduğunu ve farklı zemin koşullarında yürüyebildiğini göstermiştir. Aynı çalışmada, ödül yapısında enerjiyle ilişkili terimler bulunmasına rağmen, kol salınımının açıkça dayatılmadığı bir durumda bacaklarla ters fazlı doğal bir kol salınımı davranışının kendiliğinden ortaya çıktığı da rapor edilmiştir. Bu bulgu, kol salınımı ile yürüme performansı arasındaki ilişkinin yalnızca önceden tanımlanmış biyomekanik varsayımlarla değil, öğrenme süreci içinde ortaya çıkan davranış örüntüleri üzerinden de incelenebileceğini göstermektedir. Benzer biçimde Li vd. (2024), DRL tabanlı bir çerçeveyle yürüme, koşma, zıplama ve ayakta durma gibi farklı iki ayaklı hareket biçimlerinin tek bir yapı altında ele alınabileceğini göstermiştir. Bu sonuç, ödül fonksiyonu tasarımı ile durum geçmişinin ve durum uzayının seçiminin robot yürüyüşü başarımı açısından belirleyici önem taşıdığını ortaya koymaktadır. Ayrıca güncel derleme çalışmaları, iki ayaklı robot yürüyüşünde ödül tasarımı, gerçek sisteme aktarım, güvenilir karşılaştırma protokolleri ve bütün-vücut koordinasyonu gibi başlıkların halen temel araştırma alanları arasında yer aldığını vurgulamaktadır. Bu çerçevede, kol-bacak koordinasyonunun öğrenme davranışı, yürüme kararlılığı ve enerjiyle ilişkili ölçütler üzerindeki etkisinin ayrıntılı biçimde incelenmesi, mevcut literatürde anlamlı bir araştırma ihtiyacına karşılık gelmektedir (Radosavovic vd. 2023; Li vd. 2024; Tang vd. 2024; Bao vd. 2026).

Kaymak (2023) tarafından gerçekleştirilen çalışmada, insansı robotların yürüme becerilerinin DRL yöntemiyle geliştirilmesi ele alınmıştır. Çalışmada, geleneksel yürüme şablonu üreteci ile DRL tabanlı yaklaşımlar birlikte kullanılarak yürüyüş parametrelerinin iyileştirilmesi ve robotun yürüme kararlılığının artırılması amaçlanmıştır. Bu kapsamda Webots benzetim ortamında çift derin Q ağı tabanlı bir yöntem kullanılarak yürüyüş parametrelerinin optimize edilmesine odaklanılmış, ardından sagittal düzlemde duruş dengeleme problemi için ayrı bir yapı önerilmiştir. Tezin devamında, robotun eğimli

yüzeylerde kararlı biçimde yürüyebilmesi amacıyla PID tabanlı ve DDPG tabanlı iki farklı denetim çerçevesi geliştirilmiştir ve hem benzetim hem de gerçek robot deneyleri üzerinden bu yapıların performansları karşılaştırılmıştır. Elde edilen bulgular, önerilen DRL tabanlı çerçevenin düz zeminde daha kararlı bir yürüyüş sağladığını, eğimli yüzeylerde ise DDPG tabanlı denetleyicinin PID denetleyiciye kıyasla daha başarılı sonuçlar verdiğini göstermiştir. Bu yönüyle söz konusu tez, DRL yöntemlerinin insansı robotlarda hem yürüyüş parametrelerinin iyileştirilmesinde hem de denge denetiminde kullanılabileceğini ortaya koyan kapsamlı çalışmalardan biri olarak değerlendirilebilir.

Tüm bu çalışmalar birlikte değerlendirildiğinde, insan yürüyüşünde kol salınımının yalnızca ikincil bir hareket olmadığı; açısal momentumun düzenlenmesi, gövde rotasyonunun sınırlandırılması, enerji ekonomisi ve dış bozucu etkiler sonrasındaki toparlanma açısından işlevsel bir bileşen olduğu anlaşılmaktadır. İnsansı robotlarda ise kol salınımının etkisi, robotun morfolojisine, aktif DoF yapısına, ödül fonksiyonuna ve kullanılan denetim mimarisine bağlı olarak değişebilmektedir. Klasik model tabanlı yaklaşımlar bu ilişkiyi çoğunlukla azaltılmış modeller ve momentum dengeleme ilkeleri üzerinden açıklarken, DRL tabanlı yöntemler aynı ilişkiyi doğrudan veri temelli biçimde öğrenebilmekte ve bazı durumlarda insan yürüyüşüne benzer kol-bacak koordinasyon örüntülerini kendiliğinden ortaya çıkarabilmektedir. Bununla birlikte, farklı kol stratejilerinin yürüme kararlılığı, enerjiyle ilişkili ölçütler, öğrenme davranışı ve biyomekanik uygunluk üzerindeki etkilerinin birlikte ve sistematik biçimde ele alındığı çalışmaların sınırlı olduğu görülmektedir (Herr ve Popovic 2008; Pontzer vd. 2009; Bruijn vd. 2010; Radosavovic vd. 2023; Bao vd. 2026). Bu çerçevede, mevcut çalışmadan farklı olarak bu tezde; sadeleştirilmiş ancak 3B çok cisimli fiziksel yapıyı koruyan bir insansı model üzerinde, ES-SAC hibrit denetleyicisi altında sabit, anti-normal ve normal kol stratejileri hem düz hem engebeli zeminde sistematik biçimde karşılaştırılmış; öğrenme eğrileri, yakınsama davranışı ve çok ölçütlü kararlı durum yürüyüş analizleri birlikte sunulmuştur. Bu yaklaşım, kol salınımını yalnızca nihai yürüyüş çıktısı üzerinden değil, öğrenme süreci ve enerjiyle ilişkili davranışlar üzerindeki etkisiyle birlikte ele alan bütüncül bir karşılaştırma çerçevesi oluşturmaktadır.

3. MATERYAL VE METOT

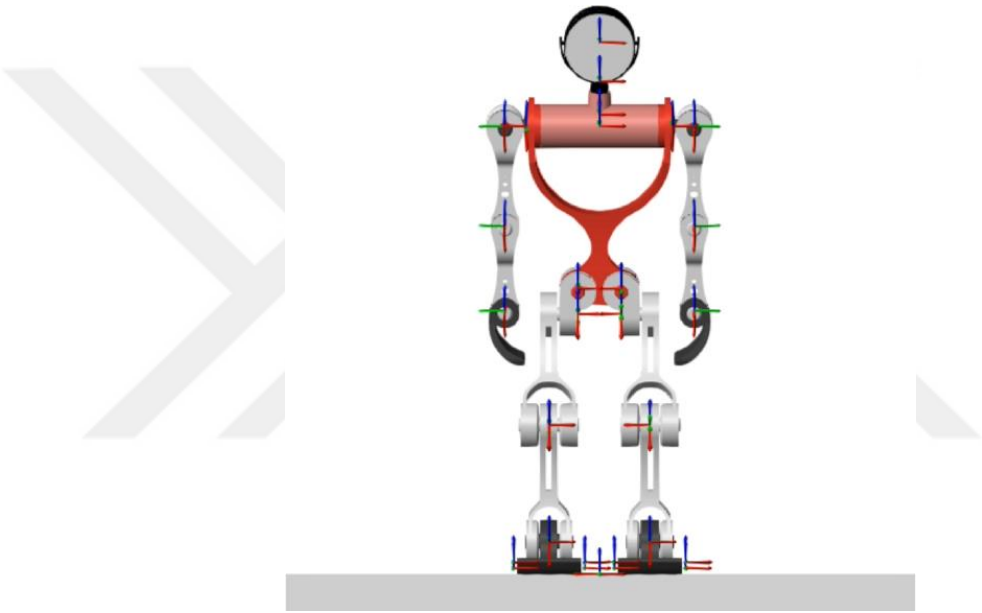
Bu bölümde tezde kullanılan insansı robot modelinin, kinematik/dinamik formülasyonların ve DRL tabanlı denetleyici yapısının hangi bileşenlerden oluştuğu ve nasıl kurulduğu sistematik olarak sunulmaktadır. Böylece sonraki sonuç ve tartışma bölümlerinde raporlanan bulguların, hangi varsayımlar, modeller, algoritmalar ve parametreler altında elde edildiği açık ve tekrar üretilebilir bir çerçeve içinde tanımlanmaktadır.

3.1. İnsansı Robot Modeli

Bu çalışmada kullanılan insansı robot modeli, insan yürüme hareketini bilgisayar ortamında inceleyebilmek amacıyla üç boyutlu uzayda tanımlanmış çok cisimli bir mekanik sistem olarak oluşturulmuştur. Burada çok cisimli sistem ifadesi, robotun gövde, bacak ve kol gibi birbirine eklemlerle bağlı birden fazla rijit parçadan oluştuğunu belirtmektedir. Bununla birlikte, modelin öğrenme sürecini gereksiz ölçüde karmaşık hale getirmemesi için eklem yapısı bilinçli olarak sadeleştirilmiştir. Başka bir ifadeyle, gerçek insan vücudunda daha fazla hareket serbestliği bulunmasına rağmen, bu çalışmada yalnızca yürüme davranışını temsil etmek için gerekli görülen temel eklem hareketleri korunmuştur. Böylece hem modelin hesaplama yükü azaltılmış hem de ES-SAC tabanlı denetleyicinin öğrenme süreci daha yönetilebilir hale getirilmiştir. Şekil 3.1'de gösterildiği üzere robotun her bir bacağı, kalça, diz ve ayak bileği olmak üzere tek eksenli üç dönel eklemden oluşmaktadır. Bu nedenle alt ekstremiteler bacak başına 3 olmak üzere toplam 6 DoF sahiptir. DoF, bir mekanik yapının bağımsız olarak gerçekleştirebildiği hareket sayısını ifade etmektedir. Bu çalışmada söz konusu alt ekstremite eklemlerinin tamamı doğrudan tork ile kontrol edilmektedir. Yani denetleyici, her bir ekleme uygulanacak döndürme etkisini belirleyerek robotun yürüme hareketini oluşturmaktadır. Üst ekstremitelerde ise her bir kol üç eklemli bir yapı şeklinde modellenmiştir. Ancak bu eklemlerin tamamı aktif olarak kontrol edilmemiştir. Modelin sade tutulması amacıyla dirsek ve bilek eklemleri sabitlenmiş, yalnızca omuz eklemleri denetim kapsamına alınmıştır. Bu nedenle kol salınımının yürüyüş üzerindeki etkisi, omuz eklemine uygulanan hareketler üzerinden incelenmiştir. Böyle bir tercih yapılmasının temel nedeni, çalışmanın odağının ayrıntılı kol biyomekaniğinden çok, kol salınımının genel yürüyüş kararlılığı ve hareket performansı üzerindeki etkisini değerlendirmek olmasıdır. Eklem eksenlerinin anatomik karşılıkları, eklem alt sistemleri içinde tanımlanan sabit dönüşüm ve hizalama blokları aracılığıyla belirlenmiştir. Bu yapı sayesinde modeldeki her eklemin hangi düzlemde ve hangi eksen etrafında hareket edeceği açık biçimde tanımlanabilmiştir. Sonuç olarak oluşturulan modelde toplam 8 adet aktif, yani doğrudan denetlenen, DoF bulunmaktadır. Bunların 6 adeti bacak eklemlerine, 2 adeti ise omuz eklemlerine karşılık gelmektedir. Bu yapılandırma, bir yandan insan yürüyüşünün temel mekanik özelliklerini temsil ederken, diğer yandan RL sürecinin uygulanabilir ve izlenebilir bir çerçevede yürütülmesine olanak sağlamaktadır.

Bu aktif DoF tanımı, modelin tüm kinematik serbestliklerini değil, doğrudan denetleyici tarafından tork komutuyla sürülen mafsalları ifade eder. Model, yere sabitlenmiş bir tabana bağlı katı bir yapı olarak değil, serbest uzayda hareket edebilen çok cisimli bir sistem olarak kurulmuştur. Oluşturulan modellemede yer (ground) dışında toplam 21 rijit gövde bulunmaktadır. Şekil 3.2'de gösterildiği gibi eklemsel hareket, bacaklarda her bir bacak için üçer adet olmak üzere kalça, diz ve ayak bileği üzerinde

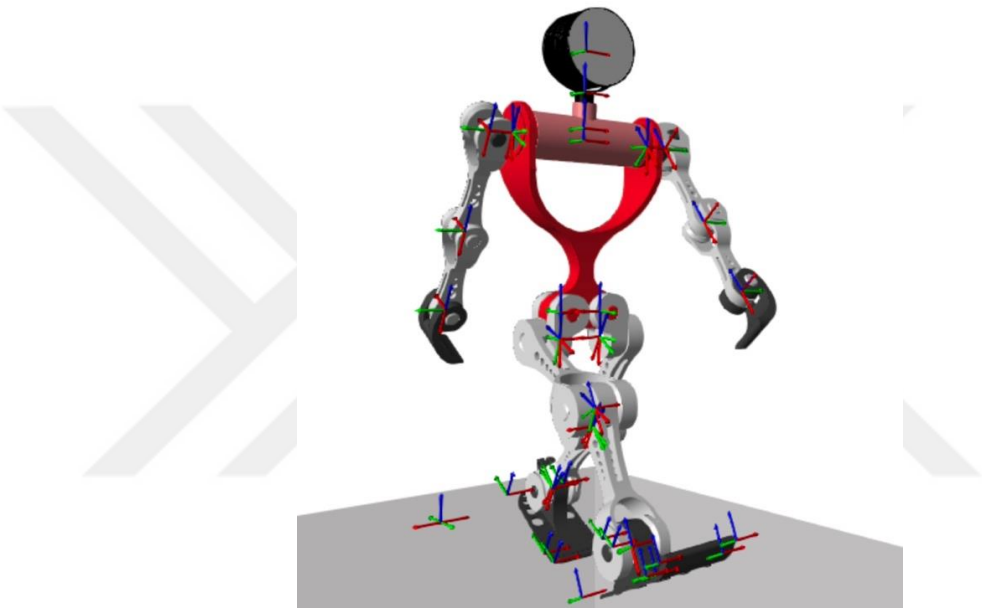
tanımlanan toplam 6 adet dönel mafsal (revolute joint) ile sağlanmaktadır. Üst ekstremite tarafında ise yalnızca omuz düzlemindeki omuz sagittal mafsalı dönel mafsal olarak tanımlanmış, omuz frontal mafsalı ise kilitli bırakılmıştır; karşı tarafta ve diğer kol segmentlerinde ayrıca dönel mafsal tanımlanmamıştır. Buna ek olarak gövdenin uzayda serbest hareketini temsil eden 6 DoF, gövdeyi bir referans çerçevesine bağlayan ve üç doğrusal (P_x, P_y, P_z) bileşenden oluşan bir burç tipi mafsal (Bushing Joint) üzerinden modellenmiştir. Dolayısıyla mekanizmanın toplam DoF sayısı, 6 (bacak dönel mafsalları) + 1 (omuz dönel mafsalı) + 6 (burç tipi mafsal) = 14 olarak değerlendirilmelidir. Bu çerçevede aktif DoF sayısının toplam DoF sayısından daha küçük raporlanması bir çelişki değildir; aktif DoF yalnızca torkla sürülen mafsalları kapsarken, kalan serbestlikler gövdenin serbest uzay hareketini ve torkla sürülmeyen bileşenleri temsil eder.



Şekil 3.1. İnsansı robot modelinin önden görünüşte eklem yerleşimleri ve koordinat eksenleri

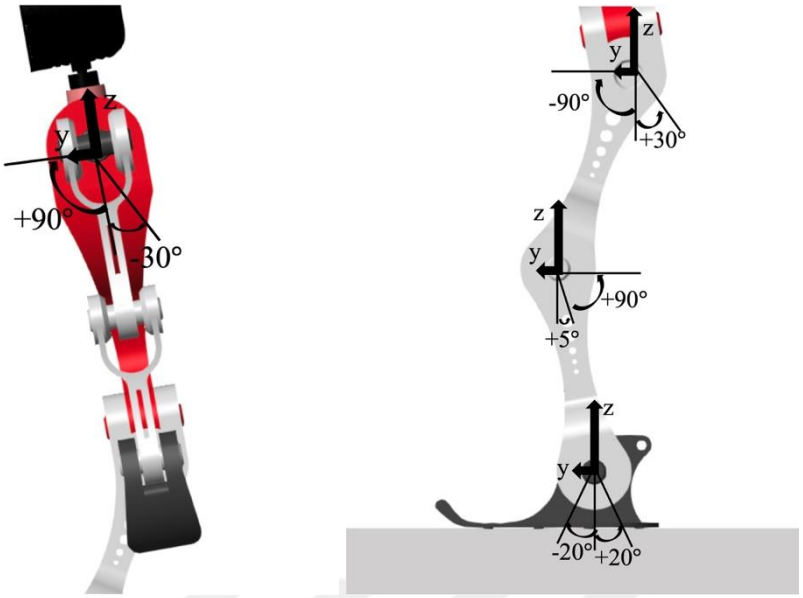
Eklem hareket açıklıkları, insan anatomisiyle tutarlı kalacak biçimde açısal sınırlar (eklem limitleri) kullanılarak kısıtlanmıştır. Şekil 3.3a'da omuz eklemi için tanımlanan açısal sınırlar, Şekil 3.3b'de ise kalça, diz ve ayak bileği eklemlerine ait açısal limitler gösterilmektedir. Dolayısıyla kalça eklemi −90° ile +30° aralığında sınırlandırılmış; böylece bacağın gövdeye göre belirlenmiş bir zarf içinde salınmasına izin verilirken aşırı açısal sapmalar engellenmiştir. Diz eklemi +5° ile +90° aralığıyla kısıtlanmıştır; sıfırdan büyük bir alt sınırın seçilmesi, dizin tam açılma konumuna ulaşmasını önleyerek temas dinamiğinde sayısal kararlılığı artırmayı ve adımdan adıma daha gerçekçi geçişler elde etmeyi amaçlamaktadır. Ayak bileği eklemi −20° ile +20° aralığında sınırlandırılmıştır; bu, ayak-yer teması sırasında plantar fleksiyon (ayağın aşağı yönde bükülmesi) ve dorsifleksiyonun (ayağın yukarı yönde bükülmesi) aşırı değerlere ulaşmasına bağlı kararsızlıkları ve temas kuvvetlerindeki ani artışları önlemeye yönelik

bir güvenlik payı sağlar (Roaas ve Dersson 1982; Soucie vd. 2011). Üst ekstremitede omuzun sagittal düzlemdeki hareketi −30° ile +90° aralığıyla sınırlandırılmıştır. Bu sınırlama, kol salınımını biyomekanik açıdan makul bir aralıkta tutar ve dinamik olarak kararsız olabilecek ya da hedeflenen kol salınımı davranışıyla uyumsuz kol duruşlarını engeller. Kol zincirindeki diğer eklemler, örneğin dirsek ve bilek, pasif ya da kilitli olarak modellendiğinden, omuz eklemine ait bu sınırlar üst ekstremitenin etkin hareket zarfını fiilen belirlemekte ve kollar için temel bir kinematik kısıt oluşturmaktadır. Sonuç olarak bu kısıtlar, modelin gerçekçi eklem hareket aralıkları içinde kalmasını sağlarken temas ve gövde dinamikleri açısından sayısal olarak kararlı bir simülasyon rejimini sürdürmeyi amaçlar; böylece ES-SAC denetleyicisinin daha hızlı öğrenmesi de desteklenmiş olur.



Şekil 3.2. İnsansı robot modelinin perspektif görünüşte eklem yerleşimleri ve koordinat eksenleri

Bu limitlere ek olarak modelde temas ve gövde dinamiklerinin sayısal kararlılığını artırmaya dönük başlangıç hedefleri ve referans tanımları da yer almaktadır. Özellikle gövdenin uzayda serbest hareketini veren bushing eklemi, simülasyonun başlangıç anında gövde konumu ve yöneliminin denetlenebilir bir başlangıç durumuna oturmasını sağlayacak biçimde hedef ve önbaşlat (prestart) değerleriyle tanımlanabilmektedir. Bu tür başlangıç koşulları, temasın ilk anda aşırı kuvvet üretmesine veya gövdenin sayısal olarak kararsız bir konfigürasyona düşmesine karşı bir güvenlik katmanı sağlar. Bu bütünleşik yapı sayesinde, modelin serbest uzaydaki gövde hareketi, ayak-yer etkileşimi ve eklem sınırları birlikte ele alınarak hem biyomekanik açıdan makul hem de sayısal olarak kararlı bir öğrenme ortamı oluşturulmuştur.



(a) (b)

Şekil 3.3. a) Üst ekstremitede omuz ekleminin açısal limitleri; **b)** Alt ekstremitede kalça, diz ve ayak bileği eklemlerinin açısal limitleri

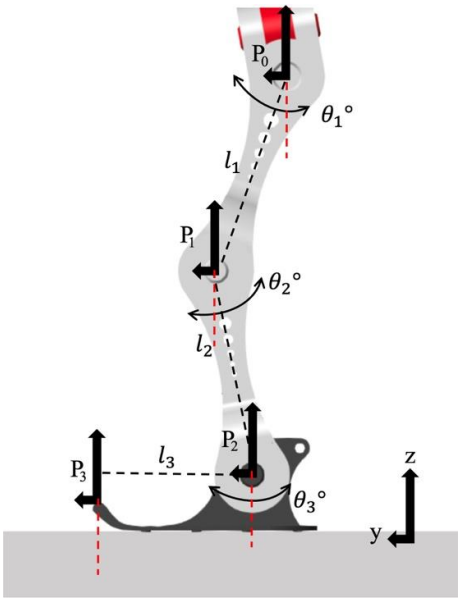
3.2. İnsansı Robot Bacağının Kinematik ve Dinamik Modeli

Bu bölümde, çalışmada kullanılan insansı robotun kinematik ve dinamik modelleme çerçevesi sunulmaktadır. Robot modeli Matlab programının Simscape Multibody ortamında 3B birçok cisimli sistem olarak kurulmuş, rijit cisimler arası bağlantılar dönel eklemlerle temsil edilmiş ve eklemler tork girdileriyle doğrudan sürülmüştür. Simülasyon tarafında sistemin hareket denklemleri, her bir rijit gövdenin kütle-atalet özellikleri, eklem kısıtları ve yerçekimi etkisi dikkate alınarak çok cisimli dinamiğin sayısal çözümü üzerinden elde edilir. Bu tür fizik tabanlı simülasyonlarda hareket, genelleştirilmiş koordinatlar cinsinden tanımlanır ve eklem/hizalama bloklarının getirdiği kinematik kısıtlar ile temas modelleri bir arada ele alınır. Sonuç olarak, robotun dinamiği kapalı formda tek bir diferansiyel denklem takımı olarak değil; kısıtlar ve temas kuvvetleri ile birlikte çözülen diferansiyel-cebirsel bir yapı içinde sayısal olarak bütünleşik biçimde hesaplanır. Bu çalışmada amaç, Simscape Multibody'nin kullandığı uzaysal çok cisimli dinamiği bütünüyle yeniden türetmekten ziyade, kontrol tasarımını ve öğrenme sürecinde anlam kazanan büyüklükleri daha açık bir matematiksel temsil ile açıklamaktır. Bu nedenle bacak kinematiği, yürüyüşün baskın bileşenini oluşturan sagittal düzlem (planar) hareketini esas alan 3 DoF bir seri zincir olarak ele alınmış; bu zincir üç dönel (revolute) eklemden oluşur ve mühendislik literatüründe hem RRR (üç dönel eklem) hem de 3R (three revolute joints) olarak adlandırılır. Her iki kısaltma da aynı yapıyı ifade eder. İleri ve ters kinematik bağıntılar ile hız kinematiği bu indirgenmiş yapı üzerinde türetilmiştir. Dinamik modelleme tarafında ise, aynı düzlemsel bacak temsili üzerinden Lagrange yaklaşımı kullanılarak indirgenmiş hareket denklemleri elde edilmekte, böylece eklem torkları ile eklem hareketleri arasındaki ilişki açık biçimde ifade edilmektedir. Bu yaklaşım, bir yandan simülasyonda korunan uzaysal temas ve

gövde dinamiği gerçekçiliğini muhafaza ederken, diğer yandan tez anlatımında kinematik ve dinamik ilişkilerin izlenebilir, denetlenebilir ve tekrar üretilebilir bir matematiksel çerçevede sunulmasına imkân verir.

3.2.1. Düzlemsel üç serbestlik dereceli bacağın kinematik modeli

Şekil 3.4'teki model, insansı bir robotun sağ bacağının sagittal (uzunlamasına) düzlemde (y-z düzleminde) yürüyüş hareketini temsil eden RRR yapılı bir kinematik zincirdir. Eklemler sırasıyla kalça, diz ve bilek eklemleridir. Kalça ile diz arasındaki uzunluk l_1, diz ile ayak bileği arasındaki uzunluk l_2, bilek ile ayak ucu arasındaki uzunluk l_3 olarak gösterilir. Her bir eklem dönme ekseni düzleminden dik (örneğin sayfa düzleminden dışarı doğru) olacak biçimde yerleştirilmiştir ve tüm eklemler aynı düzlem içinde ve yönde (aynı eksen etrafında) dönmektedir. Pozitif eklem açıları, genel mekanik yapıya uygun olarak, ilgili ekleme dışarıdan bakıldığında saat yönünün tersi yönünde dönme ile tanımlanmıştır (Spong vd. 2006; Lynch ve Park 2023). Bu durum, trigonometrik hesaplarda kosinüs bileşeninin yatay eksende, sinüs bileşeninin ise düşey eksen yönünde pozitif yer değiştirme yaratacağı anlamına gelir. Şekil 3.4'te Kalça eklemi orijin olarak alınmış ve $P_0 = (y_0, z_0) = (0,0)$ olarak tanımlanmıştır. Diz, bilek ve ayak ucu noktaları sırasıyla $P_1 = (y_1, z_1)$, $P_2 = (y_2, z_2)$ ve $P_3 = (y_3, z_3)$ ile gösterilmiştir. Kalça ekleminin bulunduğu noktanın koordinat sisteminin orijini olarak seçildiği kabul edilmektedir. Buna göre, orijinde tanımlanan koordinat sisteminde y ekseni yatay ileri yönü, z ekseni ise dikey yukarı yönü temsil etmektedir. Başlangıç konfigürasyonunda, tüm eklem açıları $\theta_1 = \theta_2 = \theta_3 = 0$ olduğunda, bacak bağlantılarının tamamının y ekseni boyunca doğrusal olarak hizalandığı varsayılmaktadır. Diğer bir ifadeyle, tüm bağlantılar yatay eksenin pozitif yönü boyunca uzanmaktadır. Eklem açıları için pozitif yön saat yönünün tersi olarak tanımlanmıştır. Dolayısıyla, örneğin θ_1 açısının pozitif olması halinde, kalça ile diz arasındaki birinci bağlantı y ekseninin pozitif doğrultusundan z ekseninin pozitif doğrultusuna doğru dönmektedir.



Şekil 3.4. Sağ bacağın sagittal düzlemdeki düzlemsel 3R kinematik modeli, eklem açıları ve bağlantı uzunlukları

Düzlemsel 3R bacak için ileri kinematik

Bu kısımda, düzlemsel 3R bacak yapısında kalça, diz ve ayak bileği eklem açıları verilerek ayak ucunun (uç efektörün) konum ve yöneliminin nasıl hesaplandığı gösterilmektedir. Böylece yürüyüş sırasında hedeflenen ayak yörüngeleri için gerekli geometrik ilişkiler tanımlanmakta ve ilerleyen hız, statik ve dinamik modellemelerin üzerine kurulacağı temel kinematik çerçeve oluşturulmaktadır.

İleri kinematik, verilen eklem açıları ($\theta_1, \theta_2, \theta_3$) için ayak ucunun (uç efektörün) konumunu ve yönelimini elde etmektir. Orijin kalça ekleminde sabitken, kalça, diz ve bilek eklemleri sırasıyla θ_1, θ_2 ve θ_3 açıları kadar dönerek bacağın yapılandırılmasını (konfigürasyonunu) belirler. Bu çalışmada tüm eklem açıları sagittal düzlemde saat yönü tersine pozitif alınmış, y ekseni ileri yönü, z ekseni yukarı yönü göstermektedir; θ_2 ikinci bağlantının birinci bağlantıya göre bağıl dönme açısıdır. Uç nokta olarak ayak ucunun orijine göre konum koordinatlarını (y, z) bulmak için bağlantı vektörlerinin yatay ve düşey bileşenleri toplanır (Spong vd. 2006; Lynch ve Park 2023).

Kalça ve diz arasındaki bağlantı

Bu kısımda, kalça ekleminden itibaren birinci bağlantı boyunca uzanan diz noktasının konum koordinatları ayrıntılı olarak ifade edilmektedir. Kalçadan itibaren θ_1 açısı kadar dönen diz noktasının, bacağın birinci bağlantı uzunluğu l_1 boyunca uzanan rijit bir çubuk olarak kabul edildiği ve bu çubuğun uç noktasının düzlemsel (y, z) koordinatlarıyla ifade edildiği göz önüne alındığında, söz konusu eklem noktasının konum koordinatları Denklem (3.1)'de verilmiştir:

$$\begin{aligned} y_1 &= l_1 \cos\theta_1 \\ z_1 &= l_1 \sin\theta_1 \end{aligned} \tag{3.1}$$

Burada, $cos\theta_1$ yatay eksen boyunca, $sin\theta_1$ dikey eksen boyunca uzunluğu gösterir.

Diz ve bilek arasındaki bağlantı

Bu bölüm, diz eklemine göre ikinci bağlantının dönmesiyle elde edilen ayak bileği konumunun matematiksel ifadesini vermektedir. İkinci bağlantı, birinci bağlantıya göre θ_2 kadar döner; dolayısıyla ikinci bağlantının mutlak yönelimi $(\theta_1 + \theta_2)$'dir. Bileğin (ayak bileği ekleminin) orijine göre konumu, diz konumuna diz-bilek bağlantı vektörü eklenerek Denklem (3.2)'deki gibi elde edilir:

$$y_2 = y_1 + l_2 \cos(\theta_1 + \theta_2)$$
$$z_2 = z_1 + l_2 \sin(\theta_1 + \theta_2) \quad (3.2)$$

Bilek ve ayak ucu arasındaki bağlantı

Bu bölümde, bilek eklemi ile ayak ucu arasındaki son bağlantı için konum ilişkileri tanımlanarak uç efektörün (ayak ucunun) global koordinat sistemine göre nihai konumu ve yönelimi elde edilmektedir. Ayak bağlantısı da bileğe göre θ_3 açısı yapar. Ayak ucunun mutlak yönelim açısı, sabit (global) koordinat sisteminin yatay eksenine göre $\theta_1 + \theta_2 + \theta_3$ olarak tanımlanır. Ayak ucunun orijinden konumu, bilek noktasına ayak bağlantısının eklenerek Denklem (3.3)'deki gibi elde edilir:

$$y_3 = y_2 + l_3 \cos(\theta_1 + \theta_2 + \theta_3)$$
$$z_3 = z_2 + l_3 \sin(\theta_1 + \theta_2 + \theta_3) \quad (3.3)$$

Elde edilen bağıntıların birleştirilmesiyle, ayak ucunun ileri kinematik konum ifadeleri Denklem (3.4)'teki biçimiyle elde edilir.

$$y(\theta_1, \theta_2, \theta_3) = l_1 \cos\theta_1 + l_2 \cos(\theta_1 + \theta_2) + l_3 \cos(\theta_1 + \theta_2 + \theta_3)$$
$$z(\theta_1, \theta_2, \theta_3) = l_1 \sin\theta_1 + l_2 \sin(\theta_1 + \theta_2) + l_3 \sin(\theta_1 + \theta_2 + \theta_3) \quad (3.4)$$

Burada, tüm bağlantıların yatay projeksiyonları toplamı y-koordinatını, düşey projeksiyonları toplamı ise z-koordinatını verir.

İleri kinematik sonucunda, uç efektörün konum ve yönelimini tanımlayan üç bağıntı elde edilir. Ayak ucunun yönelimi ise tüm eklem açıları toplamına eşit olup Denklem (3.5)'te verilmiştir:

$$\phi(\theta_1, \theta_2, \theta_3) = \theta_1 + \theta_2 + \theta_3 \quad (3.5)$$

Burada, ϕ ayak bağlantısının global yatay eksene göre açısıdır. Pozitif ϕ, ayak ucunun yere göre saat yönünün tersine döndüğünü (parmak ucunun yukarı kalktığını) ifade eder.

Sonuç olarak, elde edilen ileri kinematik bağıntılar, verilen eklem açıları için uç efektörün düzlemsel konumunu ve yönelimini tam olarak belirleyen kapalı form ifadeler

sağlamaktadır. Ancak yürüyüş denetimi açısından pratikte daha sık karşılaşılan durum, hedeflenen ayak ucu konum ve yönelimine karşılık gelen eklem değişkenlerinin tayin edilmesi, yani ters kinematik probleminin çözülmesidir. İzleyen alt bölümde, düzlemsel 3R bacak yapısı için bu ters kinematik probleminin analitik olarak nasıl formüle edildiği ve çözüldüğü ayrıntılı biçimde ele alınmaktadır.

Düzlemsel 3R bacak için ters kinematik

Ters kinematik tanım olarak robotun uç efektörünün istenen konum ve yönelimini sağlayacak eklem açılarını ($\theta_1,\ \theta_2, \theta_3$) hesaplamaktır. Düzlemsel 3R bacak için ters kinematik çözüm, ileri kinematiğe kıyasla daha karmaşıktır çünkü genel olarak bir konum için birden fazla eklem açısı çözümü bulunabilir. Ayrıca bacağın sahip olduğu DoF, istenen parametre sayısından fazla ise sonsuz sayıda çözüm mümkündür. Bu nedenle, ters kinematik çözüm iki duruma göre ele alınmalıdır. Bu nedenle ters kinematik, verilen eklem açılarına karşılık gelen uç efektör konumunu hesaplayan ileri kinematikten farklı olarak, hedef konum ve yönelimden başlayıp eklem uzayına geri dönen ters yönlü bir haritalama problemi olarak tanımlanır.

Düzlemsel 3R yapı için hedef (y, z, ϕ) verildiğinde çözüm genellikle ayrık olup iki farklı geometrik dal (diz konumunun hedef doğrultusunun iki tarafında kaldığı konfigürasyonlar) elde edilebilir; fiziksel eklem limitleri uygulamada geçerli dalı belirler.

Konum ve yönelim ile ters kinematik

Bu başlık altında, düzlemsel 3R bacak için tanımlanan ters kinematik denklemlerinin, hedef ayak ucu konumu ve yönelimi verilmişken eklem açılarını nasıl belirlediği ayrıntılı biçimde açıklanmaktadır. Böylece uç efektör uzayında tanımlanan görev hedeflerinin eklem uzayına nasıl geri yansıtıldığı, birden fazla olası diz konfigürasyonunun hangi geometrik nedenlerle ortaya çıktığı ve tekillik bölgelerinde çözümün nasıl sınırlandığı kavramsal olarak netleştirilmektedir.

Uç efektörün konumu ve yönelimi birlikte belirlendiğinde, düzlemsel 3R bacak modeli için ters kinematik çözümü Denklem (3.4) ve Denklem (3.5)'e dayalı üç skaler denklem sistemi olarak ifade edilir. Bu denklem sistemi için çözüm kümesi sürekli değil, sonlu sayıda ayrık konfigürasyondan oluşmaktadır. Düzlemsel 3R mekanizmalarda tipik olarak iki farklı çözüm elde edilmekte olup, bunlar aynı uç efektör hedefine karşılık gelen farklı diz konfigürasyonlarını temsil etmektedir. Dolayısıyla aynı ayak ucu konumu ve yönelimi için kalça, diz ve ayak bileği eklemlerine ait iki farklı eklem açısı kümesi yazılabilmektedir. Bu iki çözüm, θ_2 açısının işaretinin değişmesini değil, diz noktasının geometrik olarak hedef doğrultusunun iki tarafında konumlanması ile ortaya çıkan iki farklı dalı temsil eder; pratikte geçerli dal, fiziksel eklem limitleri ve süreklilik (bir önceki duruma yakınlık) kriteri ile seçilir. Bu tekil anda kalça, diz ve ayak bileği eklemleri doğrusal hizada olur ve mekanizma tek bir doğru üzerinde uzanır. Bu anda Jakobiyen (Jacobian) matrisi tekil olur ve dolayısıyla matrisin tersi alınamaz. Tekillik sınırında bacak, diz ileri konfigürasyonundan diz geri konfigürasyonuna, ya da tersine, süreklilik korunarak geçiş yapamaz; sistem bu pozisyonda yön değiştirerek karşı tarafa bükülemez. Eğer diz eklemi $\theta_2 > 0$ olacak biçimde sınırlandırılmışsa, tekillik durumu teorik olarak $\theta_2 \to 0$ sınırında ortaya çıkar; pratikte eklem limitleri nedeniyle tekilliğe tam olarak girilmeden çalışma aralığı korunur (Spong vd. 2006; Lynch ve Park 2023).

İstenen ayak ucu konumunu $P_d = (y_d, z_d)$ ve yönelimini (yatay eksene göre) ϕ_d olarak tanımlansın. Buradaki hedef, $\theta_1, \theta_2 \; ve \; \theta_3$ eklem açılarını analitik olarak bulmaktır. Bu amaçla, ters kinematik çözüm geometrik ilişkilerden yararlanılarak elde edilir.

Ayak bileği konumunun bulunması

Ayak bileği ekleminin hedef konumu, ayak ucunun konumundan ayak bağlantısına ait l_3 uzunluğunun geriye doğru çıkarılmasıyla elde edilir. Buna göre ayak bileği noktası, ayak ucuna göre ϕ_d doğrultusunda l_3 kadar ötelenmiş bir konumda olup bu ilişki Denklem (3.6)'da verilmiştir:

$$y_a = y_d - l_3 \cos \phi_d$$
$$z_a = z_d - l_3 \sin \phi_d \tag{3.6}$$

Burada, $y_a \; ve \; z_a$ noktaları kalça eklemine bağlı 2 bağlantılı ($l_1$ve l_2) alt zincirin ucunun ulaşması gereken konumdur. Ayak ucundan geri doğru yatay bileşen, $l_3 \cos \phi_d$ (yönü eksi yatay eksen yönünde) iken ayak ucundan geri doğru dikey bileşen ise $l_3 \sin \phi_d$ (yönü eksi dikey eksende) olur. Dolayısıyla problem, ayak ucunun hedefi (y_d, z_d, ϕ_d) yerine, bilek noktasının hedefi olan (y_a, z_a) konumuna indirgenmiş olur. Başka bir ifadeyle, ayak bileği noktası bu aşamadan itibaren alt zincirin uç noktası (uç efektör) olarak ele alınır ve ters kinematik hesabı, uzunlukları l_1 ve l_2 olan iki bağlantılı düzlemsel seri mekanizmanın, orijinden (y_a, z_a) noktasına erişecek biçimde eklem açılarını belirleme problemine dönüştürülür.

Diz eklemi açısının (θ_2) bulunması

Bu kısımda, kalça–diz–ayak bileği noktalarının oluşturduğu üçgen üzerinde kosinüs teoremi kullanılarak diz eklemi açısının analitik ifadesi elde edilmektedir. Bu üçgende kenar uzunlukları $l_1$ve l_2 olup, bilek hedef noktasının kalça eklemine olan uzaklığı $d = \sqrt{y_a^2 + z_a^2}$ şeklinde tanımlanır. Buna göre kosinüs teoremine dayalı ilişki Denklem (3.7)'de verilmiştir:

$$d^2 = l_1^2 + l_2^2 - 2\, l_1\, l_2 \cos(\pi - \theta_2) \tag{3.7}$$

Burada $(\pi - \theta_2)$ ifadesi, kalça-diz bağlantısı ile diz-ayak bileği bağlantısı arasında kalan iç açıyı temsil etmektedir. Eklem açısı θ_2 dış bükey açı yönünde tanımlandığından, üçgenin iç açısı $\pi - \theta_2$ olmaktadır.

Bu ilişkinin θ_2 cinsinden düzenlenmiş biçimi Denklem (3.8)'de verilmiştir:

$$\cos(\pi - \theta_2) = \frac{l_1^2 + l_2^2 - d^2}{2\, l_1\, l_2} \tag{3.8}$$

Denklem (3.8), $\cos(\pi - \theta_2) = -\cos\theta_2$ özdeşliği kullanılarak, Denklem (3.9)'daki biçime dönüştürülür:

$$\cos\theta_2 = \frac{y_a^2 + z_a^2 - l_1^2 - l_2^2}{2\, l_1\, l_2} \tag{3.9}$$

Buradaki ifade, diz açısının kosinüsünü yalnızca bağlantı uzunlukları ve hedef bilek konumunun kalçaya uzaklığı cinsinden yazarak ters kinematik problemini sadeleştirmektedir. Bu sayede, verilen hedef noktanın bacak uzunlukları ile erişilebilir olup olmadığı doğrudan bu eşitsizlikten kontrol edilebilir. $c = \cos\theta_2$ değerinin fiziksel olarak anlamlı olabilmesi için $-1 \leq c \leq 1$ koşulunu sağlaması gerekir. Aksi durumda, hedeflenen (y_a, z_a) noktası verilen bacak uzunlukları ile erişilebilir değildir. Bu aşamada, diz eklemi açısı θ_2 için temel çözüm Denklem (3.10)'da verilmiştir:

$$\theta_2 = \arccos\left(\frac{y_a^2 + z_a^2 - l_1^2 - l_2^2}{2\, l_1\, l_2}\right) \tag{3.10}$$

Bu çözüm, [0,π] aralığında temel çözüm olarak alınır. Geometrik olarak ortaya çıkan iki farklı çözüm dalı, θ_2'nin işaret seçiminden değil, bir sonraki adımda $\theta_1 = \alpha \pm \beta$ ilişkisinden kaynaklanmaktadır.

Kalça eklemi açısının (θ_1) hesaplanması

Kalça açısının belirlenmesinde, kalça ile ayak bileği arasındaki doğrultu ile bağlantı uzunlukları arasındaki geometrik ilişkilerden yararlanılır. Şekil 3.5'te, ters kinematik çözümünde kullanılan α ve β açılarına ilişkin geometrik tanımlar gösterilmiştir. Kalçadan bileğe çizilen doğrunun yatay eksenle yaptığı açı (α), hedef bilek konumunun yön açısını ifade etmekte olup her iki çözüm için ortak bir büyüklüktür ve Denklem (3.11)'de verilmiştir:

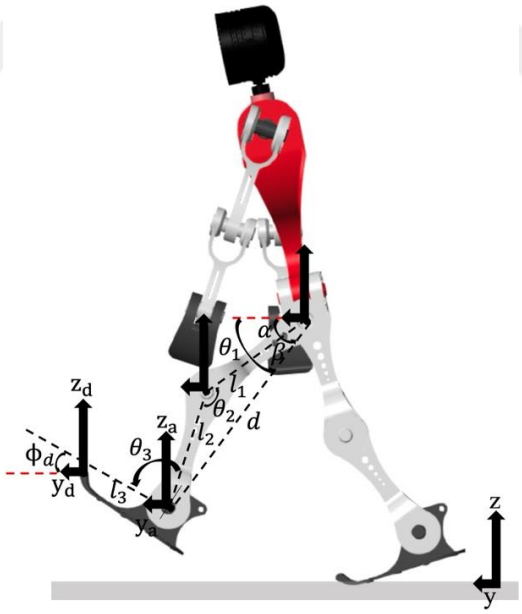
$$\alpha = \arctan 2\,(z_a, y_a) \tag{3.11}$$

Burada $arctan\,2$ fonksiyonu, $z_a\ ve\ y_a$'nın işaretini dikkate alarak doğru bölgeye ait açıyı verir.

Bir önceki adımda elde edilen iç açı β, kosinüs teoremi yardımıyla Denklem (3.12)'deki biçimde yazılabilir:

$$\beta = \arccos\left(\frac{l_1^2 + l_2^2 - d^2}{2\,l_1\,d}\right) \tag{3.12}$$

Burada β, orijin-diz-bilek üçgeninde kalça eklemindeki iç açıyı temsil etmektedir. Başka bir ifadeyle bu açı, orijin ile diz doğrultusu arasındaki eksen ile orijinden bilek noktasına çizilen doğru arasındaki açıdır. Geometrik olarak iki farklı ters kinematik çözüm dalı bulunduğundan, aynı büyüklükte ancak farklı dallara karşılık gelen iki olası konfigürasyon ortaya çıkmaktadır. Bu durum, dizin öne ya da arkaya büküldüğü iki farklı çözümle ilişkilidir.



Şekil 3.5. Sağ bacağın sagittal düzlemde ters kinematik çözümünde kullanılan geometrik açı tanımları ve bağlantı uzunlukları

Geometrik çözümleme sonucunda kalça eklemi açısı θ_1, α ve β açıları yardımıyla Denklem (3.13)'teki gibi hesaplanabilir:

$$\theta_1 = \alpha \pm \beta \tag{3.13}$$

Bu bölümde hesaplanan β açısı [0,π] aralığında tanımlıdır ve iki çözüm dalı $\theta_{1,1} = \alpha - \beta$ ve $\theta_{1,2} = \alpha + \beta$ olarak elde edilir. Seçilen dal, eklem limitleri ile uyumlu olmalı ve uygulamada süreklilik için bir önceki zamandaki çözüme en yakın olan tercih edilmelidir.

Kalçadan bileğe çizilen doğrunun yatay eksenle yaptığı açı α, hedef bilek konumunun yön açısını ifade etmekte olup her iki çözüm dalı için ortaktır ve Denklem (3.11)'de verilmiştir. Burada arctan2 fonksiyonu, z_a ve y_a bileşenlerinin işaretlerini dikkate alarak açının doğru bölgede belirlenmesini sağlar. Bir önceki adımda elde edilen iç açı β, kosinüs teoremi yardımıyla Denklem (3.12)'deki biçimde ifade edilir. Burada β, orijin-diz-bilek üçgeninde kalça eklemindeki iç açıyı temsil etmektedir. Geometrik olarak iki farklı ters kinematik çözüm dalı bulunduğundan, aynı uç noktaya karşılık gelen iki farklı konfigürasyon elde edilebilmektedir. Geometrik çözümleme sonucunda kalça eklemi açısı θ_1, α ve β açıları yardımıyla Denklem (3.13)'teki gibi hesaplanabilir. Bu bölümde hesaplanan β açısı [0,π] aralığında tanımlıdır. Buna bağlı olarak iki çözüm dalı, $\theta_{1,1} = \alpha - \beta$ ve $\theta_{1,2} = \alpha + \beta$, biçiminde elde edilir. Uygulamada seçilecek çözümün eklem limitleriyle uyumlu olması ve hareket sürekliliğinin korunabilmesi için bir önceki zaman adımındaki çözüme en yakın dalın tercih edilmesi uygun olmaktadır.

Bilek eklemi açısının (θ_3) hesaplanması

Son olarak, istenen ayak yönelimi ϕ_d'nin sağlanabilmesi için ayak bileği eklemi açısı θ_3 hesaplanır. İleri kinematikte elde edilen $\phi = \theta_1 + \theta_2 + \theta_3$ bağıntısından yararlanılarak θ_3 açısı Denklem (3.14)'te verildiği biçimde elde edilir:

$$\theta_3 = \phi_d - (\theta_1 + \theta_2) \tag{3.14}$$

$\theta_1 \; ve \; \theta_2$ açıları belirlendikten sonra, bunlara karşılık gelen θ_3 değeri Denklem (3.14) kullanılarak hesaplanır. θ_3 değerinin işareti, ayak ucunun hedef yönelime ulaşabilmesi için ayak bileği ekleminin dönme yönünü göstermektedir. Buna göre negatif değerler saat yönündeki, pozitif değerler ise saat yönünün tersindeki dönmeyi ifade etmektedir. Böylece ters kinematik çözüm tamamlanmış olur. Eğer aynı uç efektör hedefine karşılık birden fazla olası çözüm elde edilirse, uygulamada seçim eklem limitleri ve bir önceki durumla süreklilik ölçütü dikkate alınarak yapılabilir.

Konum kısıtı altında ters kinematik çözümü

Ayak ucunun yalnızca hedef konumu (y_d, z_d) verilip yönelimi serbest bırakıldığında, sistem kinematik olarak artıksıl hâle gelir. Bu durum oluştuğunda, verilen bir konum için ters kinematikte tekil bir çözüm değil, sonsuz sayıda çözümden oluşan bir çözüm eğrisi (manifold) elde edilir. Bunun nedeni, eklem sayısının uç efektörün konumunu belirlemek için gerekli DoF sayısından fazla olmasıdır. Bu durumda

çözümlerden biri genellikle serbest parametre olarak seçilir. Örneğin bilek eklem açısı (θ_3) serbest bırakılır; sonra da istenen (y_d, z_d) konum için kalça ve diz eklemleri, iki bağlantılı düzlemsel bir mekanizmanın ters kinematiği yöntemi ile hesaplanır. Bu durumda seçilen θ_3 değeri, $\phi = \theta_1 + \theta_2 + \theta_3$ bağıntısı üzerinden ayak yönelimini dolaylı olarak belirler. θ_3 açısının farklı değerleri, ayak ucunu aynı konumda tutarken, ayağın farklı açılarla (farklı yönelimlerle) durmasına izin verir. Örneğin $\theta_3 = 0°$ seçildiğinde ayak yere paralel olabilirken, $\theta_3 = 20°$ seçildiğinde ayak ucu aynı noktada kalır fakat ayağın kendisi yukarı doğru eğik bir pozisyona geçer. Bu değişimi telafi etmek için θ_1 ve θ_2 açıları uygun biçimde yeniden ayarlanır. Bu belirsizlik genellikle ek bir kısıt ile giderilir ve problem tekrar "tam belirlenmiş" durumuna getirilir. Ayağın yere paralel olması şartı ($\phi_d = 0°$) veya dizin daima ileri yöne bükülmesi (ters bükülmenin yasaklanması) gibi ilave kriterler seçilmesi durumunda artıksıllık ortadan kaldırılır ve çözüm uzayı tekillik noktaları hariç ayrık (izole) çözümlere indirgenir.

3.2.2. Düzlemsel bacak için hız kinematiği (Jakobiyen) ve statik modellemesi

Bu kısımda, önce ileri kinematikte elde edilen uç nokta konum ifadelerinin zamana göre türevleri alınarak hız kinematiği tanımlanmakta, ardından Jakobiyen matrisi kullanılarak eklem hızları ile uç nokta hızları arasındaki ilişki ortaya konmaktadır. Devamında aynı Jakobiyen yapısı, statik durumda dış kuvvetlerin eklem momentlerine nasıl yansıtıldığını göstermek için kullanılmakta ve böylece hız kinematiği ile statik kuvvet analizinin ortak bir çerçevede birleştirilmesi sağlanmaktadır.

İnsansı robot bacağının düzlemsel $y - z$ eksenlerindeki ileri kinematik yapısında, ayak ucunun konumu genelleştirilmiş eklem değişkenleri q_1, q_2 ve q_3 cinsinden tanımlanmaktadır. Bu değişkenlerden q_1, kalça ekleminin pozitif y eksenine göre ölçülen ve saat yönünün tersine pozitif kabul edilen açısını ifade etmektedir. q_2 ve q_3 ise sırasıyla diz ve ayak bileği eklemlerine ait göreli dönme değişkenleridir. Tüm eklemlerin birbirine paralel dönme eksenlerine sahip olması nedeniyle hareket yalnızca $y - z$ düzleminde gerçekleşmekte ve sistem düzlemsel bir mekanizma olarak ele alınmaktadır. Bu kabuller altında elde edilen ileri kinematik ifadeler Denklem (3.15)'te sunulmuştur:

$$\begin{aligned} y(q_1, q_2, q_3) &= l_1 \cos q_1 + l_2 \cos(q_1 + q_2) + l_3 \cos(q_1 + q_2 + q_3) \\ z(q_1, q_2, q_3) &= l_1 \sin q_1 + l_2 \sin(q_1 + q_2) + l_3 \sin(q_1 + q_2 + q_3) \end{aligned} \tag{3.15}$$

Denklem (3.15)'deki konum fonksiyonlarının zamana göre türevi alınarak uç noktanın hız ilişkisi elde edilebilir. Uç noktanın doğrusal hız bileşenleri $v_y = \dot{y}$ ve $v_z = \dot{z}$, eklem açısal hızlarına bağlıdır. Zincir kuralı kullanılarak türev alındığında her bir eklem hızının uç noktaya katkısı ortaya çıkar (Spong vd. 2006; Lynch ve Park 2023). Örneğin q_1'de küçük bir artış, l_1, l_2 ve l_3 boyunca konumu değiştirecektir. Denklem (3.15)'teki konum fonksiyonlarının zamana göre türevi alınarak uç noktanın doğrusal hız bileşenleri elde edilir ve bu ilişki Denklem (3.16)'da gösterilmiştir:

$$\begin{aligned}\dot{y} &= -l_1 \sin q_1\, \dot{q}_1 - l_2 \sin(q_1 + q_2)\,(\dot{q}_1 + \dot{q}_2) \\ &\quad - l_3 \sin(q_1 + q_2 + q_3)\,(\dot{q}_1 + \dot{q}_2 + \dot{q}_3) \\ \dot{z} &= l_1 \cos q_1\, \dot{q}_1 + l_2 \cos(q_1 + q_2)\,(\dot{q}_1 + \dot{q}_2) \\ &\quad + l_3 \cos(q_1 + q_2 + q_3)\,(\dot{q}_1 + \dot{q}_2 + \dot{q}_3)\end{aligned} \tag{3.16}$$

Denklem (3.16)'da verilen ifadeler, uç nokta hızını eklem hızları cinsinden doğrusal bir kombinasyon olarak vermektedir. Bu ilişki, matris formunda Jakobiyen matrisi ile ifade edilebilir. Jakobiyen, J(q) ile gösterilen 2×3 boyutlu bir matristir ve her bir elemanı, uç nokta konumunun ilgili eklem açısına göre kısmi türevinden oluşmaktadır. Buna göre verilen ifadeler, uç nokta hızını eklem hızlarının doğrusal birleşimi olarak tanımlamaktadır. Bu ilişki matris biçiminde yazıldığında hız dönüşümü Denklem (3.17)'deki gibi elde edilir:

$$\begin{bmatrix}\dot{y} \\ \dot{z}\end{bmatrix} = J(q) \begin{bmatrix}\dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3\end{bmatrix} \tag{3.17}$$

Jakobiyen matrisi J(q), Denklem (3.18)'de tanımlanan 2×3 boyutlu matristir:

$$J(q) = \begin{bmatrix}\dfrac{\partial y}{\partial q_1} & \dfrac{\partial y}{\partial q_2} & \dfrac{\partial y}{\partial q_3} \\ \dfrac{\partial z}{\partial q_1} & \dfrac{\partial z}{\partial q_2} & \dfrac{\partial z}{\partial q_3}\end{bmatrix} \tag{3.18}$$

Bu matris, üç eklem değişkenine karşılık gelen hızların, düzlemsel uç nokta hız vektörünü hangi doğrultularda ve hangi büyüklükte etkilediğini özetleyen kompakt bir gösterim sunmaktadır.

Denklem (3.18)'de verilen Jakobiyen matrisinin açık formu ise Denklem (3.19)'da sunulmuştur:

$$J(q) = \begin{bmatrix}-l_1\,\mathrm{s}(\rho_1) - l_2\,\mathrm{s}(\rho_2) - l_3\,\mathrm{s}(\rho_3) & -l_2\,\mathrm{s}(\rho_2) - l_3\,\mathrm{s}(\rho_3) & -l_3\,\mathrm{s}(\rho_3) \\ l_1\,\mathrm{c}(\rho_1) + l_2\mathrm{c}(\rho_2) + l_3\mathrm{c}(\rho_3) & l_2\mathrm{c}(\rho_2) + l_3\mathrm{c}(\rho_3) & l_3\mathrm{c}(\rho_3)\end{bmatrix} \tag{3.19}$$

Burada $\rho_1 = q_1$, $\rho_2 = q_1 + q_2$ ve $\rho_3 = q_1 + q_2 + q_3$ olarak verilmiştir. Bu semboller, uç-noktanın sırasıyla x, y ve z doğrultularındaki konum bileşenlerini göstermektedir. Ayrıca $s(\cdot)$ ve $c(\cdot)$ fonksiyonları sırasıyla sinüs ve kosinüs fonksiyonlarını ifade etmek üzere kısaltma olarak kullanılmıştır.

Denklem (3.19)'da verilen $J(q)$ matrisi, eklem uzayındaki hız vektörünü ($[\dot{q}_1, \dot{q}_2, \dot{q}_3]$) görev uzayındaki hız vektörüne ($[\dot{y}, \dot{z}]$) dönüştürür. Jakobiyen matrisinin her sütunu, ilgili eklem tek başına bir birim hızla döndüğünde uç noktanın kazanacağı hız bileşenlerini temsil eder. Örneğin, birinci sütundaki ifadeler sadece q_1'in hızı nedeniyle P noktasının y ve z yönlerindeki hızlarını verir. Bu nedenle J matrisinin sütunları, her eklemin hız üzerindeki etkisini gösteren vektörler olarak yorumlanabilir. Jakobiyen matrisi eklem değişkenlerinin anlık konumuna (açılarına) bağlıdır ve konfigürasyona göre değeri değişir.

Bir matrisin mertebesi, o matristeki doğrusal bağımsız satır ya da sütun vektörlerinin sayısını ifade eder. Manipülatörlerde Jakobiyen matrisinin erişebileceği en yüksek mertebeye sahip olması, ilgili konfigürasyonda görev uzayının tüm yönlerinde hız ve kuvvet iletiminin mümkün olduğu anlamına gelir. Bu çalışmada ele alınan 3R bacak modeli için J matrisi genel olarak 2 mertebelidir. Ancak bacak tamamen doğrultulduğunda ($q_2 = 0, q_3 = 0$), Jakobiyen matrisinin ikinci ve üçüncü sütunları aynı doğrultuda etkiler üretir ve bu nedenle matris tekilleşir. Böyle bir tekil konfigürasyonda bacak, bazı yönlerde uç noktaya hız ya da kuvvet aktaramaz; eşdeğer olarak, bu yönlerde sonlu bir görev uzayı hareketi veya kuvveti üretmek için çok büyük eklem hızları ya da momentleri gerekebilir (Spong vd. 2006; Lynch ve Park 2023).

Jakobiyen transpozu ile statik kuvvet ve moment ilişkisi

Jakobiyen matrisi yalnızca hız kinematiğini değil, aynı zamanda statik kuvvetler ile eklem momentleri arasındaki ilişkiyi de tanımlar. Başka bir ifadeyle, uç noktaya etkiyen bir dış kuvvetin eklemlerde nasıl momentler oluşturacağı, Jakobiyen matrisi yardımıyla hesaplanabilir. Böylece aynı Jakobiyen matrisinin bir yandan hız dönüşümü, diğer yandan da kuvvet aktarımı için kullanılabileceği ve bu iki analizin birbirini tamamlayan iki görünüm sunduğu vurgulanmaktadır. Düzlemsel 3R bacak için uç noktaya uygulanan dış kuvvet vektörü $F = \left[F_y, F_z\right]^T$ ve buna karşılık gelen eklem momentleri vektörü $\tau = [\tau_1, \tau_2, \tau_3]^T$ olmak üzere, bu ilişki Denklem (3.20)'de verilmiştir:

$$\tau = J(q)^T F \tag{3.20}$$

Burada, transpoze ifadesi kol veya bacak sisteminin uç noktasındaki kartezyen kuvvetler ile eklem uzayındaki momentler arasındaki dönüşümü sağlar. Bu sonucu türetmenin bir yolu olan sanal (δ) iş prensibi ile açıklanabilir. δ sembolü, matematiksel ve mekanik analizlerde bir büyüklüğün gerçek bir hareketi değil, sonsuz küçük mertebede varsayılan sanal değişimini göstermek için kullanılır. Uç noktadaki kuvvetin δr kadar yapacağı iş, eklemlerdeki momentlerin karşılık gelen küçük açı değişimleri δq ile yapacağı işe eşit

olmalıdır. Matematiksel olarak, $F^T \delta r = \tau^T \delta q$ olmalıdır. Ancak konum değişimi de $\delta r = J\delta q$ şeklinde Jakobiyen ile bağlı olduğundan, $F^T(J\delta q) = \tau^T \delta q$ elde edilir. Bu eşitlik herhangi bir δq için geçerli olacağından, vektörel olarak $F^T J = \tau^T$ sonucuna ulaşılır. Bu durum $\tau = J^T F$ biçiminde ifade edilir. Sonuç olarak, uç noktaya etkiyen kartezyen kuvvet vektörü F, Jakobiyen matrisinin transpozu ile eklem momentlerine dönüştürülür.

Yukarıdaki ilişki, genel durumda kuvvet-vektör iletiminin doğrusal bir haritasıdır. Jakobiyen $m \times n$ boyutunda ve m≠n olan (örneğin 2×2) bir matris olsa bile, $\tau = J^T F$ denklemi statikte geçerlidir ve verilen F için eklem momentleri tekil olarak belirlenir. Buna karşılık $n > m$ (artıksal veya fazladan DoF durum) iken, uç hızını üreten $\dot{q}$ için sonsuz çözüm bulunabilir; ayrıca çoklu temas gibi kısıtlı sistemlerde iç kuvvet dağılımı söz konusu olabilir. Örneğin 3R bacakta, uç noktaya sadece dikey bir kuvvet uygulanıyorsa, bunu dengelemek için kalça, diz ve bilek eklemleri arasında farklı moment dağılımları olabilir. Bu dağılımlar, tekil olmayan konfigürasyonlarda J^T denklemini sağlayan moment vektörleriyle ifade edilir. Bununla birlikte, J matrisinin tekil olduğu durumlarda robotun bacağı bazı yönlerde kuvvet üretemez. Bu duruma örnek olarak tamamen düz bir bacak konfigürasyonunda ($q_2 = 0 \; ve \; q_3 = 0$), yatay yönde kuvvet üretebilmek için kuramsal olarak sonsuz büyüklükte eklem momenti gerekmektedir. Bu, kinematik tekilliğin statik karşılığıdır ve ayrıca Jakobiyen mertebesinin düşmesi hem hız hem kuvvet iletiminde sistemin yeteneklerini kısıtlar (Lynch ve Park 2023).

3.2.3. Düzlemsel bacağın Lagrange tabanlı dinamik modellemesi

Bu kısımda, önceki kısımlarda elde edilen kinematik ve kütle-atalet tanımları kullanılarak düzlemsel 3R bacak için Lagrange yaklaşımına dayalı hareket denklemleri türetilmektedir. Böylece eklem torkları ile eklem ivmeleri arasındaki ilişki, kütle matrisi, Coriolis/merkezkaç ve yerçekimi terimleri üzerinden açık biçimde ifade edilmekte ve ilerleyen bölümlerde kullanılacak kontrol ve öğrenme algoritmaları için tam bir dinamik model sağlanmaktadır.

Bu çalışmada ele alınan mekanik yapı, düzlemde hareket eden üç DoF bir 3R robot bacağıdır. Sistem, kalça, diz ve ayak bileği eklemlerine karşılık gelen üç adet dönel eklemden oluşmaktadır. Önceki bölümlerde tanımlandığı üzere, bu eklemlere ait genelleştirilmiş koordinatlar sırasıyla q_1, q_2 ve q_3 ile gösterilmektedir. Bağlantı uzunlukları l_1, l_2 ve l_3, bağlantı kütleleri ise sırasıyla m_1, m_2 ve m_3 ile ifade edilmektedir. Her bir bağlantının CoM noktasının, ilgili bağlantının geometrik orta noktasında bulunduğu kabul edilmiştir. Ayrıca bağlantıların dönme eylemsizlik momentleri sırasıyla l_1, l_2 ve l_3 olarak tanımlanmıştır. Sistemin hareketi $y - z$ düzleminde incelenmekte olup, y ekseni yatay ileri yönü, z ekseni ise düşey doğrultuyu temsil etmektedir. Pozitif z yönü yukarı doğru seçildiğinden, yerçekimi ivmesi g negatif z yönünde etkimekte ve bu nedenle potansiyel enerji terimleri oluşturulurken ağırlık kuvveti aşağı yönlü olarak dikkate alınmaktadır.

Bağlantıların mutlak yönelim açıları, eklem değişkenleri cinsinden sırasıyla $\theta_1 = q_1$, $\theta_2 = q_1 + q_2$ ve $\theta_3 = q_1 + q_2 + q_3$ biçiminde ifade edilmektedir. Buna bağlı olarak, bağlantıların mutlak açısal hızları da $\dot{\theta}_1 = \dot{q}_1$, $\dot{\theta}_2 = \dot{q}_1 + \dot{q}_2$ ve $\dot{\theta}_3 = \dot{q}_1 + \dot{q}_2 + \dot{q}_3$ şeklinde elde edilir. Bu ifadeler, seri bağlı dönel eklem yapısına sahip mekanizmalarda

her bir sonraki bağlantının açısal hareketinin, kendisinden önce gelen eklem hareketlerinin birikimli etkisini içerdiğini göstermektedir.

Lagrange tabanlı dinamik modelin elde edilebilmesi için, her bir bağlantının CoM noktasına ait konum bileşenlerinin tanımlanması ve bu bileşenlerin zamana göre türevleri alınarak doğrusal hız ifadelerinin çıkarılması gerekmektedir. Bu amaçla, bağlantıların CoM noktalarının düzlemdeki konumları genelleştirilmiş koordinatlar cinsinden sırasıyla Denklem (3.21), Denklem (3.22) ve Denklem (3.23)'te ifade edilmektedir.

Birinci bağlantı, kalça eklemi etrafında sabit bir dönme merkezi çevresinde hareket etmektedir. CoM noktasının bağlantının orta noktasında bulunduğu kabul edildiğinde, birinci bağlantının CoM koordinatları Denklem (3.21)'de ifade edilir:

$$y_{c1} = \frac{l_1}{2}\cos q_1$$
$$z_{c1} = \frac{l_1}{2}\sin q_1 \tag{3.21}$$

İkinci bağlantının bir ucu diz eklemine, diğer ucu ise ayak bileği eklemine bağlıdır. CoM noktasının bağlantı uzunluğunun orta noktasında yer aldığı kabul edilmektedir. Bu nedenle, ikinci bağlantının CoM konumu birinci bağlantının uç noktasına göre tanımlanır. Diz ekleminin koordinatları $(y_1, z_1) = (l_1 \cos q_1 , l_1 \sin q_1)$ olduğundan, ikinci bağlantıya ait CoM koordinatları Denklem (3.22)'de verilmektedir:

$$y_{c2} = y_1 + \frac{l_2}{2}\cos(q_1 + q_2) = l_1 \cos q_1 + \frac{l_2}{2}\cos(q_1 + q_2)$$
$$z_{c2} = z_1 + \frac{l_2}{2}\sin(q_1 + q_2) = l_1 \sin q_1 + \frac{l_2}{2}\sin(q_1 + q_2) \tag{3.22}$$

Üçüncü bağlantının bir ucu ayak bileği eklemine bağlıdır. Bu bağlantının CoM konumunu belirlemek için, önce ikinci bağlantının uç noktasına karşılık gelen ayak bileği eklemi koordinatları dikkate alınır. Buna göre ayak bileği ekleminin koordinatları $(y_2, z_2) = (l_1 \cos q_1 + l_2 \cos(q_1 + q_2) , l_1 \sin q_1 + l_2 \sin(q_1 + q_2))$ biçiminde yazılır. CoM noktasının bağlantının orta noktasında yer aldığı kabul edildiğinde, üçüncü bağlantıya ait CoM koordinatları Denklem (3.23)'te verilmektedir:

$$y_{c3} = y_2 + \frac{l_3}{2}\cos(q_1 + q_2 + q_3)$$
$$= l_1 \cos q_1 + l_2 \cos(q_1 + q_2) + \frac{l_3}{2}\cos(q_1 + q_2 + q_3)$$
$$z_{c3} = z_2 + \frac{l_3}{2}\sin(q_1 + q_2 + q_3) \tag{3.23}$$
$$= l_1 \sin q_1 + l_2 \sin(q_1 + q_2) + \frac{l_3}{2}\sin(q_1 + q_2 + q_3)$$

Düzlemsel bacağın kinetik enerjisinin hesaplanması

Bu alt başlıkta, her bir bağlantının öteleme ve dönme hareketlerinden kaynaklanan kinetik enerji bileşenleri, ilgili CoM hızları ve açısal hızlar cinsinden ayrı ayrı hesaplanmaktadır. Daha sonra bu terimler toplanarak sistemin toplam kinetik enerjisi T elde edilmekte ve bu ifade Lagrange fonksiyonunun oluşturulmasında doğrudan kullanılmaktadır.

Sistemin toplam kinetik enerjisi T, her bir bağlantının öteleme ve dönme hareketlerinden kaynaklanan kinetik enerji bileşenlerinin toplamından oluşmaktadır. Düzlemsel hareket varsayımı altında, her bir bağlantının kinetik enerjisi, CoM noktasının doğrusal hızına bağlı öteleme kinetik enerjisi ile bağlantının CoM etrafındaki açısal hareketine bağlı dönme kinetik enerjisinin toplamı olarak ifade edilir. Birinci, ikinci ve üçüncü bağlantılara ait kinetik enerji ifadeleri sırasıyla Denklem (3.27), Denklem (3.30) ve Denklem (3.31)'de verilmiştir.

Birinci bağlantı için CoM hız büyüklüğü $v_{c1} = \sqrt{\dot{y}_{c1}^2 + \dot{z}_{c1}^2}$ ve açısal hız $\omega_1 = \dot{q}_1$ olarak tanımlandığında, birinci bağlantının kinetik enerjisi Denklem (3.24)'te ifade edilir:

$$T_1 = \frac{1}{2} m_1 v_{c1}^2 + \frac{1}{2} I_1 \omega_1^2 \tag{3.24}$$

Burada, $v_{c1}^2 = \dot{y}_{c1}^2 + \dot{z}_{c1}^2$ alınır. y_{c1} ve z_{c1} ifadelerinin zamana göre türevleri alınarak birinci bağlantının hız bileşenleri Denklem (3.25)'te ifade edilir:

$$\dot{y}_{c1} = -\frac{l_1}{2}\sin(q_1)\dot{q}_1$$
$$\dot{z}_{c1} = \frac{l_1}{2}\cos q_1 \, \dot{q}_1 \tag{3.25}$$

Bu sonuçlar kullanılarak birinci bağlantının CoM hızının karesi Denklem (3.26)'da elde edilir:

$$v_{c1}^2 = \left(\frac{l_1}{2}\dot{q}_1\right)^2 (\sin^2 q_1 + \cos^2 q_1) = \frac{l_1^2}{4}\dot{q}_1^2 \tag{3.26}$$

Buna bağlı olarak, $w_1 = \dot{q}_1$ olduğu dikkate alındığında birinci bağlantının kinetik enerjisinin son biçimi Denklem (3.27)'de verilir:

$$T_1 = \frac{1}{2}m_1\frac{l_1^2}{4}\dot{q}_1^2 + \frac{1}{2}I_1\dot{q}_1^2 = \frac{1}{2}\left(m_1\frac{l_1^2}{4} + I_1\right)\dot{q}_1^2 \tag{3.27}$$

İkinci bağlantı için CoM hızı v_{c2} ve açısal hız $\omega_2 = \dot{q}_1 + \dot{q}_2$ olarak tanımlandığında, ikinci bağlantının kinetik enerjisi Denklem (3.28)'de ifade edilir:

$$T_2 = \frac{1}{2}m_2 v_{c2}^2 + \frac{1}{2}I_2\omega_2^2 \tag{3.28}$$

Burada, $v_{c2}^2 = \dot{y}_{c2}^2 + \dot{z}_{c2}^2$ alınmakta olup, birinci ve ikinci eklemlerin birlikte etkisi nedeniyle ikinci bağlantının CoM hızının karesi Denklem (3.29)'da verilir:

$$v_{c2}^2 = l_1^2{\dot{q}_1}^2 + \frac{l_2^2}{4}(\dot{q}_1 + \dot{q}_2)^2 + l_1 l_2 \cos(q_2)\,\dot{q}_1(\dot{q}_1 + \dot{q}_2) \tag{3.29}$$

Böylece ikinci bağlantının kinetik enerjisinin son biçimi Denklem (3.30)'da ifade edilir:

$$\begin{aligned} T_2 = \frac{1}{2}m_2\left[l_1^2{\dot{q}_1}^2 + \frac{l_2^2}{4}(\dot{q}_1 + \dot{q}_2)^2 + l_1 l_2 \cos q_2\left(\dot{q}_1\dot{q}_2 + {\dot{q}_2}^2\right)\right] \\ + \frac{1}{2}I_2(\dot{q}_1 + \dot{q}_2)^2 \end{aligned} \tag{3.30}$$

Üçüncü bağlantı için CoM hızı v_{c3} ve açısal hız $\omega_3 = \dot{q}_1 + \dot{q}_2 + \dot{q}_3$ olarak tanımlandığında, üçüncü bağlantının kinetik enerjisi Denklem (3.31)'de ifade edilir:

$$T_3 = \frac{1}{2} m_3 v_{c3}^2 + \frac{1}{2} I_3 \omega_3^2 \tag{3.31}$$

Üçüncü bağlantının CoM hızı tüm eklem hareketlerinden etkilendiğinden, v_{c3}^2 ifadesi Denklem (3.32)'de verilir:

$$\begin{aligned} v_{c3}^2 = {} & l_1^2 \dot{q}_1{}^2 + l_2^2 (\dot{q}_1 + \dot{q}_2)^2 + \frac{l_3^2}{4} (\dot{q}_1 + \dot{q}_2 + \dot{q}_3)^2 \\ & + 2 l_1 l_2 \cos(q_2) \dot{q}_1 (\dot{q}_1 + \dot{q}_2) \\ & + l_1 l_3 \cos(q_2 + q_3)\, \dot{q}_1 (\dot{q}_1 + \dot{q}_2 + \dot{q}_3) \\ & + l_2 l_3 \cos(q_3)\, (\dot{q}_1 + \dot{q}_2)(\dot{q}_1 + \dot{q}_2 + \dot{q}_3) \end{aligned} \tag{3.32}$$

Bu eşitlikteki terimler fiziksel olarak merkezkaç ve Coriolis etkilerini temsil eder. T_3 ifadesi de benzer biçimde $\dot{q}_\iota$ terimlerinin karelerini ve çarpımlarını içerir. Bu doğrultuda, sistemin toplam kinetik enerjisi bağlantılara ait kinetik enerji bileşenlerinin toplanmasıyla Denklem (3.33)'te elde edilir:

$$T(q, \dot{q}) = T_1 + T_2 + T_3 = \frac{1}{2} \dot{q}^T D(q) \dot{q} \tag{3.33}$$

Burada, $D(q)$ sistemin atalet (kütle) matrisi olup eklem koordinatlarının bir fonksiyonudur. $D(q)$ matrisini, parça parça elde edilmiş kinetik enerji terimlerinin $\dot{q}_\iota \dot{q}_J$ cinsinden düzenlenmesi yoluyla elde edilir. $D(q)$, 3×3 boyutlu simetrik bir matristir.

Elde edilen $D(q)$ matris elemanları Denklem (3.34)'te verilir:

$$
\begin{aligned}
D_{11} &= m_1\frac{l_1^2}{4} + m_2\left(l_1^2 + \frac{l_2^2}{4} + l_1 l_2\cos q_2\right) \\
&\quad + m_3\left(l_1^2 + l_2^2 + \frac{l_3^2}{4} + 2l_1 l_2\cos q_2 + l_1 l_3\cos(q_2+q_3)\right. \\
&\quad \left. + l_2 l_3\cos q_3\right) + I_1 + I_2 + I_3 \\
D_{22} &= m_2\frac{l_2^2}{4} + m_3\left(l_2^2 + \frac{l_3^2}{4} + l_2 l_3\cos q_3\right) + I_2 + I_3 \\
D_{33} &= m_3\frac{l_3^2}{4} + I_3 \\
D_{12} = D_{21} &= m_2\left(\frac{l_1 l_2}{2}\cos q_2 + \frac{l_2^2}{4}\right) \\
&\quad + m_3\left(l_1 l_2\cos q_2 + \frac{l_1 l_3}{2}\cos(q_2+q_3) + l_2^2 + \frac{l_3^2}{4}\right. \\
&\quad \left. + l_2 l_3\cos q_3\right) + I_2 + I_3 \\
D_{13} = D_{31} &= m_3\left(\frac{l_1 l_3}{2}\cos(q_2+q_3) + \frac{l_2 l_3}{2}\cos q_3 + \frac{l_3^2}{4}\right) + I_3 \\
D_{23} = D_{32} &= m_3\left(\frac{l_2 l_3}{2}\cos q_3 + \frac{l_3^2}{4}\right) + I_3
\end{aligned}
\tag{3.34}
$$

Buradaki atalet matrisi $D(q)$, sistemin pozisyona bağlı kütlesel dağılımını temsil eder. Görüldüğü üzere D matrisi q_2 ve q_3 açılarına bağlı trigonometrik terimler içermektedir (özellikle $\cos q_2$, $\cos q_3$, $\cos(q_2+q_3)$). $m_i > 0$ ve $I_i > 0$ varsayımı altında $D(q)$ simetrik ve pozitif tanımlı bir matristir. Dolayısıyla bu durumda fiziksel olarak sistemin eylemsizlik özelliklerini yansıtır.

Düzlemsel 3R bacağın potansiyel enerjisinin hesaplanması

Bu alt başlıkta, her bir bağlantının kütle merkezinin yerden yüksekliği cinsinden tanımlanan yerçekimi potansiyel enerjisi ayrı ayrı hesaplanmaktadır. Elde edilen ifadeler, bir önceki alt başlıkta türetilen kinetik enerji terimleriyle birleştirilerek Lagrange fonksiyonunun $L = T - V$ oluşturulmasına ve dolayısıyla hareket denklemlerinin sistematik biçimde türetilmesine zemin hazırlamaktadır.

Sistemin toplam potansiyel enerjisi V, bağlantıların kütle merkezlerinin yerden yüksekliklerine bağlı yerçekimi potansiyel enerjilerinin toplamı olarak ifade edilir. Referans düzeyi olarak kalça ekleminin bulunduğu seviye seçildiğinde, bu noktadaki potansiyel enerji sıfır kabul edilir. Pozitif z ekseni yukarı yönlü tanımlandığından, her bir

bağlantının potansiyel enerjisi $m_i g z_{ci}$ biçiminde yazılır. Buna göre, birinci, ikinci ve üçüncü bağlantılara ait potansiyel enerji ifadeleri Denklem (3.35)'te verilir:

$$V_1 = m_1 g z_{c1} = m_1 g \frac{l_1}{2} \sin q_1$$

$$V_2 = m_2 g z_{c2} = m_2 g \left(l_1 \sin q_1 + \frac{l_2}{2} \sin(q_1 + q_2) \right) \tag{3.35}$$

$$V_3 = m_3 g z_{c3} = m_3 g \left(l_1 \sin q_1 + l_2 \sin(q_1 + q_2) + \frac{l_3}{2} \sin(q_1 + q_2 + q_3) \right)$$

Toplam potansiyel enerji V, bağlantılara ait potansiyel enerji bileşenlerinin toplamı olarak $V = V_1 + V_2 + V_3$ biçiminde yazılır. Bu ifadelerden yararlanılarak, genelleştirilmiş koordinatlar cinsinden yerçekimi terimleri $G_i(q) = \partial V / \partial q_i$ biçiminde elde edilir. Buna göre, yerçekimi vektörünün bileşenleri Denklem (3.36)'da verilir:

$$\begin{aligned} G_1(q) = \frac{\partial V}{\partial q_1} = &\left(m_1 \frac{l_1}{2} + m_2 l_1 + m_3 l_1 \right) g \cos q_1 \\ &+ \left(m_2 \frac{l_2}{2} + m_3 l_2 \right) g \cos(q_1 + q_2) \\ &+ \frac{m_3 l_3}{2} g \cos(q_1 + q_2 + q_3) \end{aligned} \tag{3.36}$$

$$G_2(q) = \frac{\partial V}{\partial q_2} = \left(m_2 \frac{l_2}{2} + m_3 l_2 \right) g \cos(q_1 + q_2) + \frac{m_3 l_3}{2} g \cos(q_1 + q_2 + q_3)$$

$$G_3(q) = \frac{\partial V}{\partial q_3} = \frac{m_3 l_3}{2} g \cos(q_1 + q_2 + q_3)$$

Elde edilen kinetik ve potansiyel enerji ifadeleri birlikte, Lagrange hareket denklemlerinin sistematik biçimde türetilebileceği eksiksiz bir enerji çerçevesi oluşturmaktadır.

Düzlemsel 3R bacak için Lagrange-Euler denklemlerinin çıkarılması

Sistemin toplam kinetik enerjisi $T(q, \dot{q})$ ve potansiyel enerjisi $V(q)$ tanımlandıktan sonra, hareket denklemleri Lagrange'ın ikinci tür denklemleri kullanılarak elde edilir. Bu kapsamda Lagrange fonksiyonu $L = T - V$ biçiminde tanımlanır. Buna bağlı olarak, her bir genelleştirilmiş koordinat için Euler-Lagrange ifadeleri Denklem (3.37)'de verilir (Schilling 1990).

Sistemin toplam kinetik enerjisi $T(q,\dot{q})$ ve toplam potansiyel enerjisi $V(q)$ tanımlandığında, hareket denklemleri Lagrange-Euler yaklaşımı kullanılarak elde edilir. Buna göre, her bir genelleştirilmiş koordinat için Euler-Lagrange ifadesi Denklem (3.37)'de verilir:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_\iota}\right) - \frac{\partial L}{\partial q_i} = \tau_i \tag{3.37}$$

Burada τ_i $(i = 1, 2, 3)$, ilgili ekleme uygulanan genelleştirilmiş torku temsil etmektedir. Bu ifade, sistemde yer alan tüm genelleştirilmiş koordinatlar için ayrı ayrı uygulanarak hareket denklemlerinin elde edilmesinde kullanılmaktadır.

Bu denklemler, her bir genelleştirilmiş koordinat için ayrı ayrı yazılıp T ve V ifadeleri yerine konularak çözüldüğünde, sistemin doğrusal olmayan ikinci mertebeden hareket denklemleri elde edilir. Elde edilen hareket ifadeleri vektör-matris biçiminde Denklem (3.38)'de özetlenir:

$$D(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau \tag{3.38}$$

Burada, $D(q)$, daha önce tanımlanan atalet matrisini, $C(q,\dot{q})\dot{q}$ terimi Coriolis ve merkezkaç etkilerini, $G(q)$ ise yerçekimi etkilerini temsil etmektedir. Böylece τ vektörü, eklemlere uygulanması gereken genelleştirilmiş torkları ifade etmektedir.

Düzlemsel 3R bacak için Coriolis ve merkezkaç terimlerinin elde edilmesi

Bu alt başlıkta, atalet matrisinin konfigürasyona bağlı yapısından kaynaklanan ve eklem hızlarıyla orantılı olan hız bağımlı dinamik terimler türetilmektedir. Sistemdeki atalet matrisinin $D(q)$ konfigürasyona bağlı yapısı nedeniyle, eklem hızlarının sıfırdan farklı olduğu durumlarda hızlara bağlı dinamik terimler ortaya çıkmaktadır. Bu terimler, Coriolis ve merkezkaç etkilerini temsil etmekte olup dinamik modelin temel bileşenlerinden birini oluşturmaktadır. Söz konusu hız bağlı terimler, atalet matrisinin elemanlarının kısmi türevlerinden yararlanılarak Christoffel sembolleri aracılığıyla elde edilir ve buna karşılık gelen ifade Denklem (3.39)'da verilir:

$$h_k(q,\dot{q}) = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{1}{2}\left(\frac{\partial D_{kj}}{\partial q_i} + \frac{\partial D_{ki}}{\partial q_j} - \frac{\partial D_{ij}}{\partial q_k}\right)\dot{q}_i\dot{q}_J \tag{3.39}$$

Bu ifade, hızlara bağlı terimleri toplayan $C(q,\dot{q})\dot{q}$ vektörünün k. bileşenini tanımlamakta olup ilgili eklem için Coriolis ve merkezkaç etkilerinin birleşik karşılığını vermektedir. Çalışmada sistem üç DoF'a sahip olduğundan, hız bağlı terimler vektörü $h = [h_1, h_2, h_3]^T$ biçiminde ifade edilmektedir. h_1 bileşeni, en fazla bağlaşım terimini içeren bileşen olupsin(q_2), sin(q_3) ve sin$(q_2 + q_3)$ ile ilişkili hız çarpımlarını kapsamaktadır. Bu bileşen içerisinde özellikle ikinci ve üçüncü bağlantıların hareketlerinden kaynaklanan Coriolis ve merkezkaç etkileri yer almaktadır. h_2 bileşeni, ikinci eklem etrafındaki göreli hareketten doğan hız bağlı terimleri içermektedir. Bu bileşende hem ikinci hem de üçüncü bağlantının dinamik katkıları etkili olmakta ve farklı eklem hızlarının çarpımlarına bağlı bağlaşım terimleri ortaya çıkmaktadır. h_3 bileşeni ise yalnızca üçüncü bağlantının hareketinden kaynaklanan terimleri içermekte ve özellikle $\sin q_3$ ile ilişkili merkezkaç etkilerini temsil etmektedir. Bu hız bağlı terimler, sistem dinamiğinde eklemler arasındaki karşılıklı etkileşimin belirlenmesinde önemli rol oynamaktadır. Merkezkaç terimleri genellikle açısal hızların kareleriyle, Coriolis terimleri ise farklı eklem hızlarının çarpımlarıyla ilişkilidir. Dolayısıyla $C(q,\dot{q})\dot{q}$ terimi, sistemin harekete bağlı bağlaşım etkilerini temsil etmektedir.

Sonuç olarak, $D(q)$, $C(q,\dot{q})\dot{q}$ ve $G(q)$ terimlerinin birlikte değerlendirilmesiyle düzlemsel 3R bacağın Lagrange-Euler tabanlı dinamik modeli tamamlanmış olmaktadır. Elde edilen dinamik model, sistemin belirli konum ve hız koşulları altındaki tork gereksinimlerinin belirlenmesinde ve benzetim tabanlı dinamik analizlerde kullanılabilmektedir.

3.2.4. Düzlemsel bacakta kinematik kısıtlar

Düzlemde hareket eden 3R yapılı insansı robot bacağı için holonomik kısıt, sistemin konum değişkenleri arasında yalnızca geometrik ilişkilere dayanan ve zaman türevlerini içermeyen bir bağıntı olarak tanımlanır. Bu bağlamda, ayak ucunun yere temas hâlinde sabit kaldığı kabulü altında, ayak ucunun düzlemdeki konumu değişmemektedir. Başka bir ifadeyle, ayak ucunun yatay ve düşey koordinatları sabit kabul edilmektedir. Sistem düzlemsel olduğundan ve tüm eklemler aynı düzlemde dönel hareket gerçekleştirdiğinden, ayak ucunun konumu eklem açılarına bağlı olarak ileri kinematik bağıntılarıyla ifade edilebilir (Denavit ve Hartenberg 1955; Fetter ve Walecka 1980; Millard ve Mombaur 2019). Buna göre, ayak ucunun yatay ve düşey doğrultulardaki konum bileşenleri Denklem (3.40)'ta verilir:

$$\begin{aligned} y_{\text{ayak}}(q_1,q_2,q_3) &= L_1\cos q_1 + L_2\cos(q_1+q_2) + L_3\cos(q_1+q_2+q_3) \\ z_{\text{ayak}}(q_1,q_2,q_3) &= L_1\sin q_1 + L_2\sin(q_1+q_2) + L_3\sin(q_1+q_2+q_3) \end{aligned} \tag{3.40}$$

Burada y_{ayak} ayak ucunun yatay doğrultudaki, z_{ayak} ise düşey doğrultudaki konumunu göstermektedir. Ayak ucunun yere sabit olduğu durumda, bu koordinatların belirli sabit değerlere eşit kalması gerekmektedir. Kalça, diz, ayak bileği ve ayak ucu arasındaki bağlantı uzunlukları sırasıyla L_1, L_2 ve L_3 ile ifade edilmektedir. Eklemlere ait

genelleştirilmiş koordinatlar q_1, q_2 ve q_3 ise sırasıyla kalça, diz ve ayak bileği eklem açılarını temsil etmektedir.

Denklem (3.40)'taki ifadeler kullanılarak, holonomik kısıt ifadeleri Denklem (3.41)'de verilir:

$$f_1(q_1, q_2, q_3) = y_{\text{ayak}}(q_1, q_2, q_3) - y_0 = 0$$
$$f_2(q_1, q_2, q_3) = z_{\text{ayak}}(q_1, q_2, q_3) - z_0 = 0 \quad (3.41)$$

Burada y_0 ve z_0, ayak ucunun yere temas ettiği noktanın sabit koordinatlarını temsil etmektedir. Bu iki holonomik kısıt ifadesi, ayak ucunun düzlemde kaymadan ve yerden ayrılmadan sabit kaldığını göstermektedir. Dolayısıyla $f_1(q) = 0$ ve $f_2(q) = 0$ koşulları sağlandığı sürece ayak ucunun konumu değişmemektedir. Bu tür holonomik kısıtlar, genelleştirilmiş koordinatlar arasında zamana açıkça bağlı olmayan geometrik bağıntılar oluşturarak sistemin etkin DoF sayısını azaltmaktadır.

Holonomik kısıtların hareket denklemlerine dahil edilmesi için Lagrange çarpanları yöntemi kullanılmaktadır. Bu amaçla, daha önce tanımlanan Lagrange fonksiyonuna her bir kısıt denklemi, ilgili Lagrange çarpanı ile çarpılarak eklenmektedir. Böylece kısıtları içeren genişletilmiş Lagrange fonksiyonu Denklem (3.42)'de verilir:

$$L'(q, \dot{q}, t) = L(q, \dot{q}, t) + \lambda_1(t) f_1(q) + \lambda_2(t) f_2(q) \quad (3.42)$$

Burada $\lambda_1(t)$ ve $\lambda_2(t)$, zamana bağlı Lagrange çarpanlarını temsil etmekte; L' ise kısıtları içeren genişletilmiş Lagrange fonksiyonu olarak tanımlanmaktadır. Bu genişletilmiş fonksiyon kullanılarak Euler-Lagrange denklemleri yazıldığında, kısıtların etkisi hareket denklemlerine doğrudan dahil edilmiş olur. Bu çerçevede $\lambda_1(t)$ ve $\lambda_2(t)$, kısıt kuvvetlerinin genelleştirilmiş koordinatlar üzerindeki etkilerini temsil etmektedir. Fiziksel olarak bu çarpanlar, ilgili holonomik kısıtların sağlanabilmesi için ayak ucunda oluşan tepki kuvvetleriyle ilişkilidir. Örneğin $f_2(q) = z_{\text{ayak}} - z_0 = 0$ kısıtı, ayak ucunun yerden ayrılmamasını ifade etmekte olup buna karşılık gelen düşey tepki kuvveti $\lambda_2(t)$ ile temsil edilmektedir. Benzer biçimde $\lambda_1(t)$ ise ayak ucunun yatay doğrultuda sabit kalmasını sağlayan kısıt kuvvetini temsil etmektedir.

Genişletilmiş Lagrange fonksiyonu için Euler-Lagrange ifadesi Denklem (3.43)'te verilir:

$$\frac{d}{dt}\left(\frac{\partial L'}{\partial \dot{q}_k}\right) - \frac{\partial L'}{\partial q_k} = 0 \tag{3.43}$$

Bu ifade açıldığında, L $(k = 1, \dots, n)$ içerisinde yer alan serbest sistem terimleri ile kısıt terimleri ayrı ayrı türevlenmektedir. Bu durumda serbest sisteme ait Euler-Lagrange denklemlerine ek olarak, kısıtlardan kaynaklanan ilave kuvvet terimleri de hareket denklemlerine dahil olmaktadır. Özellikle f_1 ve f_2 fonksiyonlarının q_k'ya göre kısmi türevleri, kısıtların genelleştirilmiş koordinatlar üzerindeki etkisini belirleyen terimler olarak denklemlere katılmaktadır.

Buna göre, kısıtları içeren genişletilmiş Euler-Lagrange ifadeleri $(k = 1, \dots, n)$ Denklem (3.44)'te ifade edilir:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} + \lambda_1 \frac{\partial f_1}{\partial q_k} + \lambda_2 \frac{\partial f_2}{\partial q_k} = 0 \tag{3.44}$$

Bu ifade, her bir genelleştirilmiş koordinat için kısıt kuvvetlerini de içeren hareket denklemlerini temsil etmektedir. Burada $\frac{\partial f_i}{\partial q_k}$ terimi, kısıt fonksiyonunun ilgili genelleştirilmiş koordinata göre değişimini, başka bir ifadeyle kısıt Jakobiyen'in bir elemanını göstermektedir. λ_1 ve λ_2 ise bu doğrultulardaki kısıt kuvvetlerinin büyüklüklerini belirleyen Lagrange çarpanlarıdır. Sisteme eklem torkları ya da başka tür dış genelleştirilmiş kuvvetler uygulanması durumunda, bu etkiler hareket denklemlerinin sağ tarafında ayrı bir terim olarak gösterilebilir.

Denklem (3.44)'deki ifadeler kullanılarak, $M(q)$ sistemin atalet matrisini, $C(q, \dot{q})\dot{q}$ Coriolis ve merkezkaç terimlerini, $G(q)$ yerçekimi etkilerini ve $Q = [Q_1, Q_2, Q_3]^T$ harici genelleştirilmiş kuvvetler vektörünü temsil edecek biçimde tanımlandığında, kısıt ifadelerine ait Jakobiyen matrisi Denklem (3.45)'te verilir:

$$J(q) = \begin{bmatrix} \frac{\partial f_1}{\partial q_1} & \frac{\partial f_1}{\partial q_2} & \frac{\partial f_1}{\partial q_3} \\ \frac{\partial f_2}{\partial q_1} & \frac{\partial f_2}{\partial q_2} & \frac{\partial f_2}{\partial q_3} \end{bmatrix} = \left[\frac{\partial f_i}{\partial q_j}\right]_{\substack{i=1,2 \\ j=1,2,3}} \tag{3.45}$$

Burada $J(q)$, ayak ucunun sabit konum kısıtlarının genelleştirilmiş koordinatlara göre kısmi türevlerinden oluşan Jakobiyen matrisidir. Bu matris, iki kısıt denklemi ve üç genelleştirilmiş koordinat bulunduğundan 2×3 boyutundadır. Lagrange çarpanları vektörü ise $\lambda = [\lambda_1, \lambda_2]^T$ biçiminde tanımlanmaktadır.

Denklem (3.45)'deki ifadeler kullanılarak, kısıtlı sistemin hareket ifadeleri Denklem (3.46)'da ifade edilir:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = Q - J^T(q)\lambda \tag{3.46}$$

Burada yer alan $J^T(q)\lambda$ terimi, kısıt kuvvetlerinin genelleştirilmiş koordinatlar yönündeki bileşenlerini temsil etmektedir. Bu ifade, daha önce elde edilen $\lambda_1 \partial f_1 / \partial q_k + \lambda_2 \partial f_2 / \partial q_k$ terimlerinin vektör biçimindeki karşılığıdır. Eşitliğin sağ tarafında Q harici kuvvetler vektörü yer alırken, J^T(q)λ terimi kısıtlardan doğan reaksiyon kuvvetlerinin sisteme etkisini genelleştirilmiş kuvvet formunda göstermektedir. Böylece ayak ucunun sabitliğini sağlayan kısıt kuvvetleri hareket denklemlerinde temsil edilmiş olur.

Burada, ileri ve ters kinematik bağıntılarla eklem açıları ile ayak ucu konumu arasındaki ilişki tanımlanmış; Jakobiyen matrisi üzerinden hız kinematiği ve statik kuvvet iletimi ortaya konmuştur. Lagrange yaklaşımıyla türetilen hareket denklemleri, atalet, Coriolis/merkezkaç ve yerçekimi terimlerini tek bir çerçevede bir araya getirmiş; destek fazındaki holonomik temas kısıtlarının eklenmesiyle model bütünlüğe kavuşturulmuştur. Elde edilen bu mekanik model, robotun fiziksel davranışını tam olarak tanımlayan matematiksel altyapıyı sağlamaktadır. Bundan sonraki bölümde konu, bu model üzerinde çalışacak denetleyiciye taşınmakta ve ES-SAC algoritmasının mimarisi ve bileşenleri ele alınmaktadır.

3.3. Evrimsel Strateji Destekli Yumuşak Aktör-Eleştirmen Modeli

Önceki bölümlerde insansı robot bacağının kinematik, dinamik ve kısıtlı hareket modeli ayrıntılı biçimde oluşturularak sistemin fiziksel davranışı matematiksel olarak tanımlanmıştır. Bu bölümde, söz konusu modele uygulanacak kontrol sinyalini oluşturacak olan denetleyicinin yapısı ele alınmakta ve çalışmanın temel katkısını oluşturan ES-SAC mimarisi tanıtılmaktadır. İlk olarak ES bileşeninin popülasyon tabanlı arama mekanizması açıklanmakta, ardından yumuşak aktör-eleştirmen algoritmasının politika ve değer ağlarına ilişkin yapılandırması ortaya konulmaktadır. Son olarak aktör ve eleştirmen ağ mimarileri ile bunların birlikte oluşturduğu yumuşak politika yapısı ayrıntılandırılarak, ES-SAC ajanının insansı robot yürüyüşünü nasıl öğrendiği için gerekli kavramsal çerçeve tamamlanmaktadır.

3.3.1. Evrimsel strateji algoritmasının mimarisi

ES, EA'lar ailesine ait, popülasyon tabanlı bir optimizasyon tekniğidir ve özellikle sürekli karar uzaylarında yüksek performans göstermektedir (Wang vd. 2023). ES algoritması, birden fazla aday çözümü (birey) jenerasyonlar boyunca değerlendirir, en başarılı olanları seçer ve bu bireyler üzerinde mutasyon uygulayarak bir sonraki jenerasyonun popülasyonunu üretir (Emmerich ve Shir 2018). Böylece politika parametreleri, jenerasyonlar boyunca giderek daha yüksek ödül veren bölgelere doğru

yönlendirilir. Her birey ortamda koşturulup elde edilen toplam ödül/return (dönüş) kaydedilir; bu dönüş jenerasyonun uygunluk ölçütüdür.

Bu çalışmada ES, SAC ajanının aktör ağı parametreleri üzerinde çalışmaktadır. Başlangıçta politika için rastgele parametreler atanır ve bu parametre vektörü θ ile gösterilir. Her jenerasyonda, mevcut politika parametrelerinden Gauss gürültüsü eklenerek yeni bireyler üretilir. Bu süreç, aktör ağının parametreleri etrafında tanımlanan çok değişkenli bir Gauss dağılımından örnekleme yapmak şeklinde de düşünülebilir. Bu sebeple θ, dağılımın ortalamasını (mean) temsil eder ve mevcut SAC aktörünün referans parametre vektörünü oluşturur. Her bir popülasyon bireyi ise bu ortalama etrafında, σ ile ölçeklenen rastgele bir pertürbasyonla elde edilen küçük bir varyasyonudur. Elde edilen bireyler, ortamda çalıştırılarak ödül fonksiyonu ile değerlendirilir ve uygunluk değerleri hesaplanır. Ardından elit seçimi uygulanır; en yüksek uygunluğa sahip bireyler seçilerek politika parametrelerinin güncellenmesinde kullanılır. Bu süreç tekrarlanarak yeni jenerasyonlar oluşturulur ve çözüm giderek rafine edilir (Salimans vd. 2017).

Bu başlık altında kullanılan θ sembolü, aktör ağının ağırlık ve yanlılık değerlerinden oluşan parametre vektörünü ifade etmekte olup Bölüm 3.2'deki kinematik analizde aynı sembolle gösterilen eklem açılarından (θ_1, θ_2, θ_3) farklı bir büyüklüğü temsil etmektedir. Her iki gösterim de ilgili alanlardaki yaygın konvansiyona uygun biçimde korunmuştur; bağlam ayrımı, alt indisler ve kullanım yeri üzerinden sağlanmaktadır. Popülasyonun her bir bireyinin parametreleri θ_i ile gösterilir. ES yapısının çekirdek bileşenlerinden biri olan popülasyon, çözüm uzayı içinde rastgele başlatılmış ve her jenerasyonda güncellenen bu bireylerden oluşur. Mutasyon işlemi, mevcut politika parametrelerine Gauss gürültüsü eklenmesiyle yeni bireyler üretir. ES literatüründe yaygın olarak kullanılan mutasyon adımı bu çalışmada da Denklem (3.47)'de verildiği biçimde modellenmiştir:

$$\theta_i' = \theta + \sigma\,\varepsilon_i \tag{3.47}$$

Burada θ, bir önceki jenerasyondan elde edilen aktör ağı parametre vektörünü, σ mutasyon ölçeğini ve ε_i ise $\mathcal{N}(0, I)$ dağılımından örneklenen rastgele pertürbasyon vektörünü temsil etmektedir. Bu sayede her jenerasyonda, mevcut politika parametrelerinin çevresinde dağılan yeni bireyler üretilmekte ve çözüm uzayının farklı bölgeleri araştırılabilmektedir (Obuchowicz 2019).

Seçim aşamasında, popülasyondaki en iyi bireylerin belirlenmesi amaçlanmaktadır. Salimans vd. (2017) ve Wierstra vd. (2014) çalışmalarında önerildiği üzere, her bireyin uygunluk değerlendirmesi birden fazla bölüm üzerinden elde edilen ortalama ödüle göre yapılabilmektedir. Bu çalışmada kullanılan formülasyonda, bir bireyin ortalama ödülüne karşılık gelen uygunluk değeri F_i Denklem (3.48)'de tanımlanmıştır:

$$F_i = \frac{1}{K}\sum_{k=1}^{K} R(\theta_i', k) \tag{3.48}$$

Burada F_i, i. bireyin ortalama ödülünü, yani uygunluk (fitness) değerini temsil etmektedir. K, değerlendirme sayısını; $R(\theta_i', k)$ ise θ_i parametrelerine sahip bireyin k. koşudan elde ettiği toplam ödülü göstermektedir (Wierstra vd. 2014). Bir bireyin uygunluğunun yalnızca tek bir bölüm üzerinden değerlendirilmesi, başlangıç koşullarındaki rastlantısallık ve ortamın stokastik etkileri nedeniyle yanıltıcı sonuçlar doğurabilmektedir. Bu nedenle her birey için birden fazla bağımsız değerlendirme gerçekleştirilmekte ve elde edilen toplam ödüllerin ortalaması alınmaktadır. Böylece ölçüm gürültüsü ile stokastik çevre etkileri azaltılarak her politikanın gerçek performansını daha güvenilir biçimde temsil eden bir uygunluk kestirimi elde edilmektedir.

Güncelleme adımında, yeni jenerasyona ait politika parametreleri seçilen elit bireyler üzerinden belirlenmektedir. Bu amaçla, en yüksek ortalama ödüle sahip N_e adet elit birey seçilmekte ve bu bireylere ait parametreler, uygunluk değerlerine bağlı ağırlıklar kullanılarak birleştirilmektedir. Buna göre, yeni politika parametre vektörü θ , ağırlıklı ortalama biçiminde Denklem (3.49)'te tanımlanmıştır:

$$\theta \leftarrow \sum_{j=1}^{N_e} w_j \; \theta_j' \tag{3.49}$$

Burada w_j katsayıları, j. elit bireyin yeni politika güncellemesine katkısını belirleyen ağırlıkları temsil etmektedir. Bu ağırlıkların $\sum_{j=1}^{N_e} w_j = 1$ koşulunu sağlaması, elde edilen yeni parametre vektörünün elit bireylerin konveks birleşimi olarak kalmasını güvence altına almaktadır. Ayrıca $w_j \geq 0$ koşulu, her bir elit bireyin güncellemeye negatif olmayan katkı vermesini sağlamaktadır. Bu çerçevede, yeni politika parametre vektörü elit bireylerin performans bilgilerini bir araya getiren kararlı bir güncelleme mekanizmasıyla elde edilmektedir. N_e, seçilen elit birey sayısını; θ_j', j. elit bireyin parametre vektörünü ve w_j ise ilgili bireye atanan normalize edilmiş ağırlığı temsil etmektedir. Ağırlıklar, bireylerin göreli uygunluk değerlerine bağlı olarak belirlenmekte ve daha yüksek ödül elde eden bireylerin yeni politika parametreleri üzerinde daha fazla etkili olmasını sağlamaktadır. Bu güncelleme mekanizması, Gauss dağılımının ortalamasını, yani temel politika parametre vektörünü, elit bireylerin yoğunlaştığı bölgeye doğru taşımaktadır. Böylece bir sonraki jenerasyonda gerçekleştirilen örnekleme işlemi yeni merkez etrafında yapıldığından, arama süreci giderek daha yüksek uygunluk değerine sahip çözümlerin bulunduğu bölgeye yönelmektedir. Ayrıca bu güncelleme yapısı, türev bilgisi gerektirmeyen örnekleme tabanlı optimizasyon yaklaşımıyla ilişkili

olup ES yapıları literatüründe yaygın biçimde kullanılmaktadır (Hansen ve Ostermeier 2001).

ES algoritması, keşif ve sömürü dengesini standart sapma parametresi σ değerinin dinamik olarak ayarlanması yoluyla sağlamaktadır. Arama adımının çok hızlı küçülmesi, algoritmanın yerel minimumlara erken yakınsamasına neden olabileceğinden, σ jenerasyonlar ilerledikçe kontrollü biçimde azaltılan bir adım büyüklüğü olarak ele alınmaktadır. Buna göre, standart sapma güncelleme kuralı Denklem (3.50)'te tanımlanmıştır:

$$\sigma \leftarrow (1 - \lambda)\,\sigma + \lambda\,\sigma_{min} \tag{3.50}$$

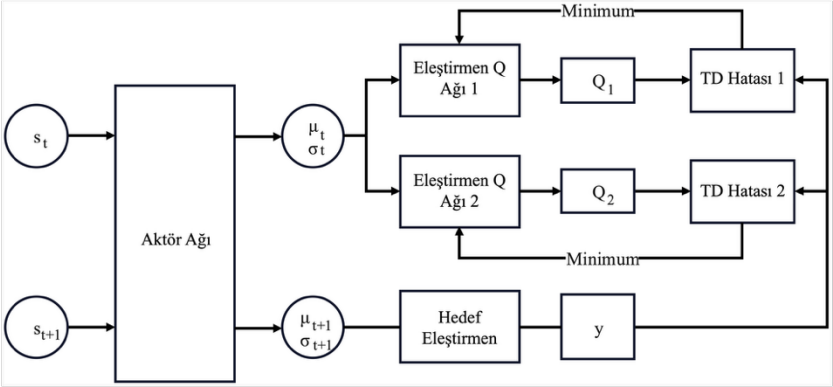
Burada λ, standart sapma azaltma katsayısını, σ_{min} ise izin verilen en küçük standart sapma değerini temsil etmektedir. Bu güncelleme yapısı, aramanın başlangıç aşamasında daha geniş bir keşif yapılmasına, ilerleyen jenerasyonlarda ise daha dar bir parametre bölgesinde daha ince ayarlı bir arama yürütülmesine olanak sağlamaktadır. Böylece eğitim sürecinin başında daha büyük σ değerleri ile çözüm uzayının geniş bölgeleri araştırılırken, sonraki aşamalarda σ'nın kademeli olarak küçülmesiyle arama süreci yüksek performanslı çözümler çevresinde yoğunlaşmaktadır.

Bu çalışmada ES modülü, SAC ajanı ile birlikte ES-RL çerçevesi içinde kullanılmaktadır. Her jenerasyonda, aktör ağının güncel parametre vektörü θ etrafında bir popülasyon oluşturulur ve bu aktör bireyler ortamda bölüm oynayarak deneyim üretir. Bu popülasyondaki her birey, SAC aktör ağının parametrelerinin bozunuma uğratılmış küçük bir varyasyonu olarak düşünülebilir. Dolayısıyla aynı jenerasyon içinde paralel olarak değerlendirilen politikalar, ortak bir temel aktörün çevresinde farklılaştırılmış örnekler sunar ve ES adımı bu örnekler üzerinden politika aramasını sürdürür. Elde edilen durum-eylem-ödül geçişleri, ortak bir tekrar oynatma belleğine eklenir. SAC algoritması, bu tampondaki örnekleri kullanarak aktör ve çift eleştirmen (double critic) ağlarının parametrelerini gradyan temelli olarak günceller; aynı zamanda entropi katsayısı da politikadaki belirsizlik düzeyine göre uyarlanır (Haarnoja vd. 2019; Lehmann 2024; Liu 2024). Böylece ES, politika parametrelerini türevsiz olarak kaba bir biçimde yönlendirirken, SAC'in gradyan temelli güncellemeleri daha ince ayarlı iyileştirmeler sağlar. Bu hibrit yapı, CEM-RL ve ES-RL literatüründe önerilen yaklaşımlarla uyumlu olup, sürekli uzaylarda zorlu yürüme görevleri için güçlü bir öğrenme performansı sunmaktadır.

3.3.2. Yumuşak aktör-eleştirmen algoritmasının mimarisi

SAC algoritması, maksimum entropi ilkesine dayanan çağdaş bir DRL yöntemidir. Geleneksel RL yaklaşımlarında amaç, verilen bir politika $\pi(a_t \mid s_t)$ için beklenen kümülatif ödülü maksimize etmektir; burada π politika fonksiyonunu, a_t zaman t'de seçilen eylemi ve s_t ise o andaki durumu göstermektedir. SAC algoritması ise bu amaca ek olarak politika dağılımının entropisini de maksimize eder. Böylece keşif (exploration) düzeyi yüksek tutulur ve daha sağlam, dengeli bir öğrenme süreci elde edilir

(Haarnoja vd. 2019). Bu prensipleri somutlaştıran SAC ajanının genel mimarisi Şekil 3.6'da verilmiştir.



Şekil 3.6. SAC ajanının genel mimarisi ve aktör-eleştirmen bileşenleri (Zhang vd. 2024)

SAC çerçevesi; tek bir aktör ağı, birbirinden bağımsız iki eleştirmen (critic) Q-ağı, bu ağların hedef (target) karşılıkları ve bir deneyim yeniden oynatma arabelleği (experience replay buffer); tahmin edilen Q-değeri ile hedef Q-değeri arasındaki farkı gösteren zamansal fark hatası (Temporal Difference, TD, error) terimlerinden oluşur. Ajan mevcut durumu s_t aldığında, aktör ağ tarafından üretilen stokastik politika Denklem (3.51)'te verildiği biçimde tanımlanmaktadır:

$$\pi_\theta(a_t \mid s_t) = \mathcal{N}\big(\mu_\theta(s_t), \sigma_\theta(s_t)\big) \tag{3.51}$$

Burada θ, aktör (politika) ağının öğrenilebilir parametrelerini (eğitim sırasında güncellenen ağ ağırlıkları ve bias terimleri) göstermektedir. $\mu_\theta(s_t)$ ve $\sigma_\theta(s_t)$, verilen s_t durumu için eylem dağılımının sırasıyla ortalama ve standart sapma değerlerini üreten fonksiyonlardır. Standart sapma $\sigma_\theta(s_t)$, politikanın ürettiği eylemlerin değişkenliğini belirler ve dolayısıyla keşif-sömürü dengesini doğrudan etkiler. $\mathcal{N}(\cdot)$ normal (Gauss) dağılımını, yani politikanın eylemleri örneklediği olasılık dağılımını ifade etmektedir (Haarnoja vd. 2018a). Böylece $\mu_t = \mu_\theta(s_t)$ ve $\sigma_t = \sigma_\theta(s_t)$ değerleri durum-bağımlı olarak öğrenilir; politika deterministik değil, içsel olarak stokastiktir ve bu da doğal bir keşif mekanizması sağlar.

Eğitim sürecinin başlangıcında eleştirmen Q-ağları $Q_k(s, a; \phi_k)$ (burada k=1,2 her bir eleştirmeni gösterir) rastgele parametrelerle başlatılır. Hedef eleştirmen ağlarının parametreleri ise başlangıçta ana eleştirmen ağlarının parametreleriyle aynı seçilmektedir. Buna göre, hedef eleştirmen parametrelerinin başlangıç eşitliği Denklem (3.52)'da verilir:

$$\phi_{tk} = \phi_k \tag{3.52}$$

Burada ϕ_k, k. eleştirmen Q-ağının öğrenilebilir parametreler vektörünü, ϕ_{tk} ise ilgili hedef (target) ağın parametrelerini göstermektedir. Benzer biçimde, aktör ağı $\pi(a \mid s; \theta)$ de rastgele parametrelerle başlatılmaktadır.

Bu temel ağ yapıları başlatıldıktan sonra bir ısınma (warm-start) fazı uygulanır. Bu aşamada ajan, rastgele bir politika izleyerek ortamda eylemler seçer ve elde edilen deneyimler (durum, eylem, ödül ve bir sonraki durum $[s, a, r, s']$) deneyim havuzunda depolanır (Srinivas vd. 2020). Eğitim safhasında, politika $\pi(a \mid s; \theta)$ ve gözlemler kullanılarak sürekli bir eylem uzayına uygun olacak biçimde sürekli değerli eylemler üretilir. Seçilen eylem ortama uygulanır, ortamdan bir ödül r ve yeni bir durum s' alınır. Elde edilen (s, a, r, s') dörtlemesi deneyim havuzuna kaydedilir (Haarnoja vd. 2018a).

Modeli eğitmek için mini-yığınlar (mini-batch) yöntemi kullanılır. Deneyim havuzunda tutulan geniş veri kümesinden rasgele seçilmiş küçük yığınlar (mini-batch'ler) çekilir. Tekrarlamalı NN yapısına dayalı sistemlerde her mini-batch, farklı deneyim dizilerinden oluşur ve her dizi ardışık çok sayıda deneyim çiftini içerir. Bu yapı, hem zamansal bağımlılıkların yakalanmasına katkı sağlar hem de öğrenme sürecinin daha kararlı ve verimli olmasına yardımcı olur (Mnih vd. 2015). Her mini-batch için, aşağıda açıklanan biçimde eleştirmen ağları, aktör ağı ve entropi katsayısı güncellenir.

İkiz eleştirmen Q-ağlarının kullanılmasının temel amacı, değer tabanlı RL'de sık karşılaşılan aşırı tahmin yanlılığını azaltmaktır. Her bir eleştirmen ağ, verilen durum-eylem çiftine karşılık gelen Q-değerini bağımsız olarak tahmin etmektedir. Daha kararlı ve güvenilir bir öğrenme süreci sağlamak amacıyla, hedef Q-değeri iki eleştirmen tahmininin minimumu kullanılarak Denklem (3.53)'de tanımlanmıştır:

$$y_i = r_i + \gamma \min_{k \in \{1,2\}} Q_{tk}(s_i', a_i'; \phi_{tk}) - \alpha \ln \pi(a_i' \mid s_i'; \theta) \tag{3.53}$$

Burada r_i, ajan tarafından alınan anlık ödülü; $\gamma \in (0,1)$ iskonto katsayısını; α ise entropi katsayısını temsil etmektedir. İskonto katsayısı ile çarpılan ve eleştirmen Q-ağlarından elde edilen minimum değer, gelecekteki ödüllerin önemini yansıtır; minimumun alınması ise aşırı tahmin yanlılığını azaltır. Alt indis terimi olan t, Q_{tk} ve ϕ_{tk} terimlerinin hedef ağlara ait olduğunu belirtir. Hedef Q-değeri, bir sonraki durum s_i' ve bu durumda politikanın ürettiği eylem a_i' kullanılarak hesaplanır (Haarnoja, Zhou, vd. 2018). Entropi terimi $\alpha \ln \pi(a_i' \mid s_i'; \theta)$, doğrudan amaç fonksiyonuna entegre edilerek öğrenme sırasında keşfi teşvik eder (Haarnoja vd. 2019). Böylece ajan hem beklenen ödülü maksimize etmeye hem de yüksek entropiye sahip, güvenilir politika çıkarımlarına ulaşmaya yönlendirilir.

Her bir eleştirmen ağ için kullanılan kayıp fonksiyonu, hedef Q-değeri ile tahmin edilen Q-değeri arasındaki fark temel alınarak Denklem (3.54)'de tanımlanmıştır:

$$L_k = \frac{1}{2M}\sum_{i=1}^{M}\left(y_i - Q_k(s_i, a_i; \phi_k)\right)^2 \tag{3.54}$$

Burada M, mini-batch boyutunu; $Q_k(s_i, a_i; \phi_k)$ ise k. eleştirmen ağın tahmin ettiği Q-değerini göstermektedir. Bu kayıp fonksiyonunun amacı, hedef değer y_i ile tahmin edilen Q-değeri arasındaki farkın karesini minimize ederek eleştirmen ağ parametrelerini güncellemek ve daha isabetli değer tahminleri elde etmektir (Haarnoja vd. 2019).

Aktör ağını güncellemek için kullanılan politika kaybı Denklem (3.55)'da tanımlanmıştır:

$$\mathcal{J}_\pi = \frac{1}{M}\sum_{i=1}^{M}\left(-\min_{k\in\{1,2\}} Q_k\left(s_i, a_i; \phi_k\right) + \alpha \ln \pi\left(a_i \mid s_i; \theta\right)\right) \tag{3.55}$$

Bu ifade, entropi düzenlemesi içeren politika optimizasyonunu temsil etmektedir. İlk terim, politikanın yüksek Q-değerine sahip eylemleri tercih etmesini sağlarken, ikinci terim politika entropisini teşvik etmektedir. Böylece hem keşif davranışı hem de yüksek ödüllü eylemlerin seçimi eş zamanlı olarak desteklenmektedir (Haarnoja vd. 2019).

Entropi katsayısı α değerinin otomatik olarak ayarlanması için kullanılan kayıp fonksiyonu Denklem (3.56)'da tanımlanmıştır:

$$L_\alpha = \frac{1}{M}\sum_{i=1}^{M}\left(-\alpha \ln \pi\left(a_i \mid s_i; \theta\right) - \alpha\mathcal{H}\right) \tag{3.56}$$

Burada $\mathcal{H}$, keşif ve sömürü arasındaki istenen dengeyi belirleyen hedef entropi değerini temsil etmektedir. Bu kaybın minimizasyonu, α değerini dinamik olarak güncellenerek politika entropisinin hedef değere yakın tutulmasını sağlamaktadır. Böylece keşif düzeyi ile ödül optimizasyonu arasında etkili bir denge korunmaktadır (Haarnoja vd. 2018a; 2019).

SAC algoritmasında eleştirmen Q-ağları, zamansal fark hatasını minimize edecek biçimde eğitilirken, aktör ağı entropi düzenlemeli politika kaybı ile güncellenmektedir (Zhang ve Xie 2024). Buna ek olarak, hedef Q-ağlarının parametreleri, eleştirmen

parametrelerinin yavaş hareket eden ortalaması alınarak güncellenmektedir. Bu güncelleme, Polyak ortalaması olarak bilinen yaklaşım temelinde gerçekleştirilmektedir. Hedef Q-ağlarının parametre güncellemesi, Polyak ortalaması yaklaşımı temelinde Denklem (3.57)'de tanımlanmıştır:

$$\phi_{tk} \leftarrow \tau\phi_k + (1-\tau)\phi_{tk}, \quad k = 1,2 \tag{3.57}$$

Burada $\tau \in (0,1]$, hedef ağ yumuşatma katsayısını temsil etmektedir. Bu güncelleme mekanizması, hedef Q-ağlarının ana eleştirmen ağlarını aniden takip etmesini önleyerek daha yumuşak ve kararlı parametre geçişleri sağlamaktadır. Böylece güncellemelerdeki dalgalanmalar azaltılmakta ve zamansal fark hedeflerinin daha kararlı biçimde hesaplanmasına katkı sunulmaktadır (Fujimoto vd. 2018; Haarnoja vd. 2018a).

Yumuşak aktör-eleştirmen politika mimarisi

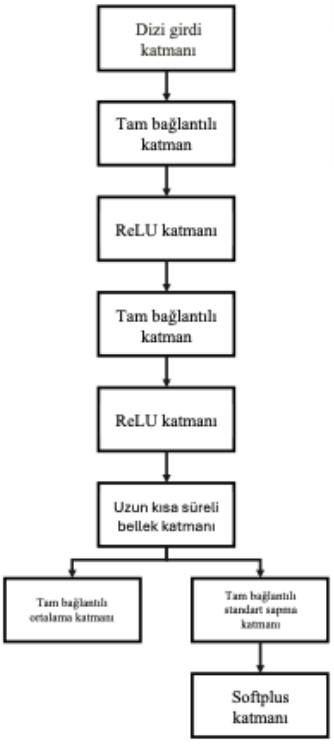
Bu çalışmada ajanın gözlem vektörü, kullanılan konfigürasyona bağlı olarak koşula duyarlı biçimde tanımlanmıştır. Normal ve anti-normal konfigürasyonlarda gözlem vektörü $o_t \in \mathbb{R}^{37}$ olup iki bileşene ayrılır: robotun anlık durumunu temsil eden $s_t \in \mathbb{R}^{29}$ ve bir önceki kontrol adımında uygulanan tork komutlarını içeren $u_{t-1} \in \mathbb{R}^{8}$. Bu ayrıştırmanın motivasyonu, özellikle temas güdümlü yürümenin yüksek dinamikliği ve eyleyici geri besleme gecikmelerine duyarlılığı nedeniyle, ajanın girişini Markov varsayımına daha yakın hale getirmektir. Sabit kol konfigürasyonunda ise gözlem vektörü daha düşük boyutludur ve $o_t \in \mathbb{R}^{31}$ olarak kullanılmıştır. Burada torkların ardışık adımlar boyunca sürekliliğini yakalamak amacıyla bir örnek gecikmeli bellek terimi (unit delay, z^{-1}) gözlem bileşenlerine dahil edilmiştir. Sensör tabanlı durum değişkenleri içinde sağ ve sol ayak temas göstergeleri, ajanın yürüyüş fazlarını ayırt etmesine olanak tanırken; gövde yönelim bileşenleri ve açısal hızlar denge korunumu ve düşüşe gidebilecek dönel kararsızlıkların erken sezimi açısından kritik bir geri besleme sağlar. Gövdenin yanal konumu, düşey yer değiştirmesi ve doğrusal hız bileşenleri ise ileri ilerleme, yanal sapma ve yer teması kaynaklı düşey salınımları doğrudan kodlar; eklem açıları ve açısal hızlar da (kalça, diz, ayak bileği ve omuzlar) adım geometrisini ve kol salınımının gövde dinamiklerine etkisini aynı gözlem uzayında değerlendirmeyi mümkün kılar.

SAC algoritmasının tekrarlayan (recurrent) varyantında, kısmi gözlemlenebilirlik problemlerini ele almak amacıyla ağ mimarisine Uzun Kısa Süreli Bellek (Long Short-Term Memory, LSTM) katmanı eklenmiştir (Hausknecht ve Stone, 2015). Bu yaklaşım, orijinal SAC (Haarnoja vd., 2019) algoritmasının ileri beslemeli ağ yapısını koruyarak çıkış katmanından önce bir LSTM modülü yerleştirmektedir.

Aktör ağı mimarisi

Bu çalışmada kullanılan aktör ağı, Şekil 3.7'de şematik olarak gösterilmiştir ve politikanın temel bileşenlerinden birini oluşturur. Aktör ağı, anlık durumu ve yürüyüşün zamansal dinamiklerini birlikte temsil edebilmek üzere tasarlanmıştır. Ağ girdisi, konfigürasyona bağlı tanımlanan gözlem vektörüdür ve normal ile anti-normal

durumlarda 37 boyutlu giriş sinyali eklem açıları, açısal hızlar ve sistemin yürüyüşe ilişkin kinematik değişkenlerini kapsar. Bu girdiler önce ikişer adet 256 nöronlu tam bağlantılı (Fully Connected, FC) katman üzerinden işlenir ve her katmandan sonra ReLU aktivasyonu uygulanarak ham gözlemlerden yürüyüş davranışını açıklayan daha soyut öznitelikler çıkarılır. Elde edilen temsil, 256 gizli birimli bir LSTM katmanına aktarılır; böylece aktör yalnızca mevcut anlık gözlemi değil, yakın geçmişteki adımlara ait bağlamsal bilgiyi de dikkate alarak eylem üretir. LSTM çıkışı iki ayrı kola ayrılarak Gauss politika parametreleri üretilir; bir kol eylem ortalamasını, diğer kol ise eylem dağılımının belirsizliğini temsil eden standart sapma terimini üretir. Bu yapı, politikanın hem kararlı hem de keşfe açık biçimde stokastik davranmasını sağlar. Ağ, girdi olarak ajanın mevcut durum vektörünü almakta ve çıktı olarak bir eylem üretmektedir. Eylem seçimi, parametreleri aktör ağı tarafından belirlenen bir olasılık dağılımından örnekleme yoluyla yapılmaktadır. DNN kullanılması sayesinde aktör ağı, durumlar ile eylemler arasındaki karmaşık ilişkileri öğrenebilmektedir (Sutton ve Barto 2018). Ağ yapısı çok katmanlı ileri beslemeli (feedforward) bir ağ veya yinelemeli (recurrent) bir ağ şeklinde tasarlanabilir (Hashemi vd. 2023). Her katmanda, ağırlık ve bias parametreleri ile doğrusal bir dönüşüm uygulanmakta, ardından aktivasyon fonksiyonu ile doğrusal olmayan bir çıktı elde edilmektedir (Sun vd. 2024).



Şekil 3.7. Çalışmada kullanılan aktör ağının katman mimarisi

Ağın ilk katmanı, ortamdan gelen durum vektörünü alarak bunu daha yüksek boyutlu bir gizli gösterime dönüştürmekte ve ham girdiden anlamlı öznitelikler

çıkarmaktadır. Ardışık FC gizli katmanlar, politikanın durum-eylem ilişkisinin ifade gücünü artırmakta ve daha zengin bir temsil sağlamaktadır. Her FC katmanından sonra kullanılan aktivasyon fonksiyonu, Doğrultulmuş Lineer Birim (Rectified Linear Unit, ReLU) fonksiyonudur. ReLU, doğrusal olmayan karar sınırlarının öğrenilmesine imkân tanır. ReLU'ya uygulanan girdi pozitif ise doğrudan çıktıya aktarılır; negatif ise sıfıra eşitlenir. Bu özellik, pozitif aktivasyonların türevlerini kaybetmeden derin katmanlardan geçebilmesini sağlar ve sigmoid aktivasyon fonksiyonlarına kıyasla kaybolan gradyan (vanishing-gradient) problemini azaltır (Goodfellow vd. 2016).

Aktör ağı, ileri beslemeli katmanlara ek olarak bir LSTM katmanı da içermektedir. LSTM katmanı, zaman boyutunun önemli olduğu veya gözlemlerin tam olmadığı (başka bir ifadeyle, sistemin kısmi gözlemlenebilir olduğu) durumlarda özellikle yararlıdır. LSTM, önceki zaman adımlarına ait girdi ve durum bilgilerini hafızasında tutarak, bu bilgiyi mevcut çıktıya dâhil edebilir. Böylece aktör, eylemleri tek bir anlık gözleme değil, zaman boyunca edinilmiş gözlemler dizisine dayanarak üretebilir ve bu sayede kaybolan gradyan sorunları hafifletilir. Bu özellikleri sayesinde, doğrudan ölçülmeyen durum bileşenleri LSTM katmanı içinde dolaylı biçimde temsil edilebilir ve tam Markov olmayan yapı kısmen telafi edilir (Bakker 2002; Icarte vd. 2019). LSTM katmanının parametreleri, zaman üzerinde zincir kuralının uygulanmasıyla gerçekleştirilen zaman boyunca geri yayılım (backpropagation through time) yöntemiyle güncellenir. Bu yaklaşım, zaman boyutunda tutarlı bir gradyan akışı sağlayarak öğrenme sürecinin kararlı olmasına katkıda bulunur (Bakker 2007).

LSTM katmanından sonra aktör ağı iki ayrı çıkış dalına ayrılmaktadır. Bu dallardan biri eylem dağılımının ortalama vektörünü, diğeri ise standart sapma değerlerini üretmektedir. Ortalama çıkışı politikanın tercih ettiği eylemleri belirlerken, standart sapma çıkışı belirsizlik düzeyini ve keşif davranışını düzenlemektedir. Standart sapma teriminin fiziksel olarak negatif değer alamaması gerektiğinden, bu dalın sonunda Softplus aktivasyon fonksiyonu kullanılmaktadır. Buna göre, Softplus fonksiyonu Denklem (3.58)'de tanımlanmıştır.

$$\mathrm{Softplus}(x) = \ln(1 + e^{x}) \tag{3.58}$$

Bu fonksiyon her zaman pozitif çıktı üretmekte ve sıfıra yakın bölgede türevi sıfıra düşmediği için gradyan bilgisinin korunmasına katkı sağlamaktadır. Böylece standart sapma parametreleri daha düzgün ve kararlı biçimde öğrenilebilmektedir. Bu yapı, ağ çıktısının deterministik bir eylem yerine parametrik bir olasılık dağılımını temsil etmesi açısından önem taşımaktadır. Bu sayede stokastik politika, olasılık yoğunluğu üzerinden türevlenebilir hâle gelmekte ve etkin biçimde optimize edilebilmektedir.

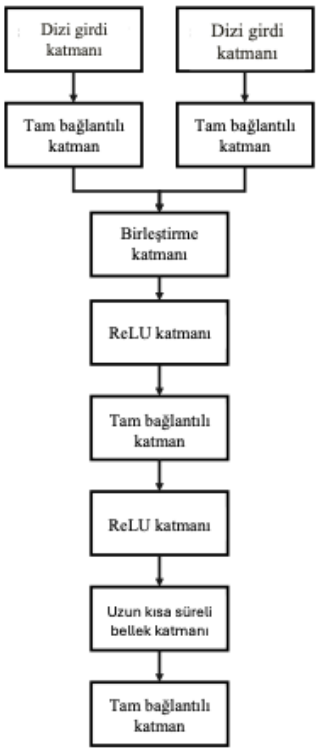
Politika entropisi yüksek olduğunda, kayıp fonksiyonundaki entropi terimi ve öğrenilen sıcaklık parametresi, ağın daha büyük standart sapma değerleri üretmesini teşvik eder. Böylece dağılım genişler ve keşif artar. Politika entropisi düşük olduğunda ise ağ belirli eylemler etrafında yoğunlaşmak istediğinde standart sapma küçülür, dağılım

daralır ve eylem seçimi daha kararlı hâle gelir. Bu keşif ve sömürü dengesi, entropi ağırlıklandırması ile ayarlanır (Zhu vd. 2024).

Öğrenme sürecinde, aktörün çıktısı olan olasılık dağılımı hem bu sürecin bir sonucu hem de belirleyici bir unsurudur. Başlangıçta SAC ajanı, yüksek entropili bir politika izleyerek geniş ölçüde keşif yapmakta ve farklı eylemleri deneyerek çeşitli deneyimler toplamaktadır. Öğrenme ilerledikçe, ajan daha yüksek ödül getiren eylemleri tercih etmeye başlar. Bununla birlikte SAC algoritmasında entropi zamanla sıfıra zorlanmaz. Bunun yerine, hedeflenen entropi düzeyine ulaşmak için sıcaklık parametresi α otomatik olarak ayarlanır (Haarnoja vd. 2019). Ajan belirli eylemlere aşırı odaklanıp keşfi azalttığında, entropi bonusu yeniden daha geniş bir keşif yapmasını teşvik eder. Bu mekanizma, ajanın yerel maksimumlara sıkışmasını engeller ve uzun vadede keşif yapmaya devam etmesini sağlar (Liu 2024). Başka bir ifadeyle SAC algoritması, keşif ve sömürüyü sistematik biçimde dengeleyerek, ajanı hem en iyi eylemleri öğrenmeye hem de gerektiğinde yeni seçenekleri denemeye yönlendirir (Lehmann 2024).

Eleştirmen ağı mimarisi

RL çerçevesinde aktör ve eleştirmen ağları eşgüdümlü olarak öğrenir. Aktör ağı, mevcut durum vektöründen yola çıkarak bir eylem önerirken, Şekil 3.8'te gösterilen eleştirmen ağı bu durum-eylem ikilisini sayısal olarak değerlendirerek bir Q-değeri üretir. Bu değer, ilgili eylemin ne kadar uygun olduğuna dair nicel bir ölçüttür. Aktör ağı, SAC algoritmasında, beklenen Q-değeri ile politika entropisini birlikte maksimize edecek biçimde güncellenir; böylece ajan hem yüksek ödüllü eylemleri tercih etmeyi öğrenir hem de keşif yeteneğini korur (Zhang ve Chen 2023).



Şekil 3.8. Çalışmada kullanılan eleştirmen ağının katman mimarisi

Kullanılan eleştirmen mimarisinde, durum ve eylem girdileri ayrı dallar üzerinden işlenmektedir. Bu çalışmada eleştirmen ağı, durum ve eylemi birlikte değerlendirerek $Q(s,a)$ kestirimi yapan iki girişli bir yapıdadır. Durum girdisi 37 boyutlu durum vektörü olup 256 nöronlu bir FC katmanla işlenerek bir durum öznitelik temsiline dönüştürülür. Aktörden gelen 8 boyutlu eylem girdisi de benzer biçimde 256 nöronlu bir FC katmandan geçirilerek eylem temsiline haritalanır. Bu iki temsil, birleştirme katmanında bir araya getirilerek durum ve eylem bilgisini ortak bir özellik uzayında birleştiren birleşik bir gösterim oluşturur. Birleştirilmiş özellikler ek bir FC ve ReLU bloğundan sonra 256 gizli birimli LSTM katmanına aktarılır; böylece değer kestirimi, yalnızca anlık duruma değil, yakın geçmiş bağlamına da duyarlı hale gelir. Son olarak tek nöronlu çıkış katmanı, ilgili durum-eylem çifti için skaler $Q(s,a)$ değerini üretir ve bu çıktı aktör güncellemelerini yönlendiren temel öğrenme sinyalini sağlar. Ortamdan gelen durum dizisi, bir dizi girdi katmanı (sequence input layer) aracılığıyla ağa alınır ve FC katmanlar üzerinden yüksek boyutlu bir durum öznitelik vektörüne dönüştürülür. Benzer biçimde, eylem girdi dizisi de ikinci bir dizi girdi katmanı ve bir FC katmanı üzerinden ayrı bir eylem öznitelik vektörüne haritalanır. Bu iki vektör daha sonra birleştirme (concatenation) katmanında birleştirilir ve ReLU aktivasyon fonksiyonu içeren doğrusal olmayan katmanlardan geçirilir. Böylece, durum-eylem uzayındaki doğrusal olmayan ilişkileri temsil eden birleşik bir gizli gösterim elde edilir.

Birleştirilmiş gizli vektör, ilave bir FC + ReLU bloğundan sonra LSTM katmanına aktarılır. LSTM katmanı, hem mevcut hem de geçmiş zaman adımlarına ait bilgiye erişerek, zaman bağımlılıklarını ve gecikmeli etkileri modelleyebilmekte ve böylece kısmi gözlemlenebilir durumlarda daha tutarlı değer tahminleri üretebilmektedir (Bakker 2007). LSTM çıkışı, son olarak tek nöronlu bir FC katmanına uygulanarak skaler bir $Q_{\phi}(s,a)$ değeri üretilir; burada ϕ, critic ağının parametrelerini temsil etmektedir.

Eleştirmen ağı için kayıp fonksiyonunun ağ ağırlıklarına göre türevleri geri yayılım (backpropagation) ile hesaplanır ve parametreler buna göre güncellenir. Optimizasyon, Uyarlamalı Moment Tahmini (Adaptive Moment Estimation, Adam) algoritmasının bir varyantı ile gerçekleştirilmiştir (Reyad vd. 2023). Adam, Stokastik Gradyan İnişi (SGD) ve RMSProp yöntemlerinin avantajlarını birleştirerek, gradyanların ve bunların karelerinin hareketli ortalamalarını kullanır ve bu sayede daha kararlı ve verimli bir güncelleme mekanizması sunar (Ilboudo vd. 2022).

Her güncelleme adımında, tekrarlayan deneyimlerden oluşan yeniden oynatma belleğinden rastgele örneklenen geçişler mini-yığınlara (mini-batch) aktarılır (Haarnoja vd. 2019). Eleştirmen ağının çıktısı olan tahmini ile hedef değer arasındaki fark, karesel hata türü bir kayıp fonksiyonu üzerinden minimize edilir. Uygun koşullar altında, bu süreç sonucunda öğrenilen fonksiyonunun yumuşak Bellman denklemine yakınsayacağı beklenmektedir. Bu işlem, SAC algoritmasında bulunan iki ayrı eleştirmen ağı için de benzer biçimde, fakat bağımsız parametre kümeleriyle yürütülür; her iki ağ da aynı hedef değeri kullanmakta, dolayısıyla öğrenme süreçleri birbirine paralel ilerlemektedir (Haarnoja vd. 2018b).

SAC ajanı, değer tahminlerindeki aşırı iyimserliği (overestimation bias) azaltmak amacıyla ikiz/çift eleştirmen (double Q) stratejisini kullanır. Bu yaklaşımda, hedef değer

hesaplanırken iki eleştirmen ağının çıktılarından minimumu alınır. Eleştirmen ağlarının güncellenmesinde hedef ağ (target network) yapıları kullanılmaktadır. Hedef ağlar, mimari olarak ilgili eleştirmen ağlarıyla özdeştir fakat parametreleri doğrudan gradyan adımlarıyla değil, Polyak ortalaması (yumuşak güncelleme) ile yenilenir. Bu sayede hedef Q-ağı, ana eleştirmen ağını bir miktar gecikmeli izleyerek daha dengeli bir referans işlevi görür ve özellikle eğitimin erken safhalarında eleştirmen ağının çıktılarındaki ani dalgalanmaları sönümler. Her eleştirmen ağı için ayrı bir hedef ağı bulunmaktadır; bu hedef ağlar, SAC eğitim sürecinin kararlılığında kritik rol oynar (Haarnoja vd. 2018b).

Bu çalışmada eleştirmen ağları, hedef ağlara kıyasla daha sık güncellenmekte; aktör ağı ise, eleştirmen ağının değer tahminlerinin yeterince olgunlaşmasına imkân verecek biçimde daha seyrek güncellenmektedir. Ayrıca politikanın stokastik yapısı (Gauss dağılımlı eylem örnekleme), eylem uzayında doğal bir yumuşatma etkisi sağlayarak aktör ağının gereksiz ve gürültülü güncellemeler yapmasını engellemeye yardımcı olur (Shen 2024). Böylece, aktör ve eleştirmen ağları birlikte, hem keşif-sömürü dengesini koruyan hem de sayısal olarak kararlı bir öğrenme süreci oluşturur.

3.4. Yumuşak Aktör-Eleştirmen Ödül Fonksiyonu

Bu çalışmada insansı robotun yürüyüşünü denetlemek için tasarlanan ödül fonksiyonu, çok bileşenli bir yapıya sahiptir ve bacak güdümlü yürüme performansını merkeze alır. Buna bağlı olarak, ileri yönde ilerlemeyi ve yürüyüş sürekliliğini önceliklendirecek biçimde formüle edilmiştir. İleri hız ve dengeli, dik duruşun sürdürüldüğü süre, yüksek öncelikli ödül terimleri olarak tanımlanmıştır. Gövdenin yanal sapması ve yer düzlemine doğru düşey alçalma güçlü biçimde cezalandırılır. Ek bir ceza terimi ise enerji verimliliğini teşvik eder. Bu yolla politika, kararlı ileri yönlü yürüyüşü maksimize etmeye yönlendirilir (Heess vd. 2017). Buna ek olarak, kol salınımını biyomekanik açıdan daha doğal hâle getirmek için iki adet kol-hareketi ödül bileşeni entegre edilmiştir. Bu kol terimlerine, yürüme performansını bozmayacak biçimde dengeyi iyileştiren ince ayar işlevi görmeleri için küçük ağırlıklar atanmıştır. Sonuç olarak ödül fonksiyonu, bacak merkezli kararlı bir insansı yürüyüşü korurken aynı zamanda insan benzeri kol-bacak koordinasyonunu da teşvik eder. Toplam ödül, farklı alt bileşenlerin ağırlıklı toplamı olarak tanımlanır. Denklem (3.59), ödül fonksiyonunun bileşenlerini göstermektedir:

$$r = w_d r_d + w_f r_f + w_l r_l + w_v r_v + w_p r_p + w_{a1} r_{a1} + w_{a2} r_{a2} \tag{3.59}$$

Burada r_d, süre (duration) ödülüdür. Her zaman adımında, robot dik konumunu koruduğu sürece sabit bir pozitif ödül verilir. r_f, ileri hareket ödülüdür. r_l, yanal sapma cezasıdır; sol-sağ (x ekseni) sapmaları en aza indirmek için uygulanan sayısal bir ceza terimidir. r_v, düşey hareket cezasıdır ve düşey (z) eksen boyunca salınımları bastırmak üzere tasarlanmıştır. r_p, güç tüketimi cezasıdır. r_{a1}, zıt taraftaki kol ve bacak arasında doğal bir ters-faz (counter-phase) salınımını teşvik eder. r_{a2}, kol simetrisi cezasıdır ve iki kolun ters-fazda simetrik biçimde hareket etmesini teşvik eder. Ödüldeki ağırlık katsayıları w ile gösterilmiştir. Her ağırlık, ilişkili olduğu ödül bileşeniyle aynı alt indisi taşımaktadır.

Sabit-kol (Fixed-arms) durumunda, kola ilişkin terimler ödül fonksiyonundan çıkarılır ve omuz eklemleri sıfır omuz torkları ile kilitlenir; dolayısıyla eğitim ve değerlendirme sırasında herhangi bir kol hareketi bulunmaz.

Denklem 3.59'daki genel ödül ifadesi, her bir alt bileşenin ağırlıklı toplamı biçiminde tanımlanmıştır. Çizelge 3.3'te yer alan üç koşula özgü ödül fonksiyonları, bu genel formun ağırlık katsayıları uygulanmış ve açılmış hâlidir.

Çizelge 3.1. Ödül/ceza terimlerinin değer aralıkları

Ödül/ceza terimleri	Değer aralığı
Süre ödülü	0.025 (sabit)
İleri hız ödülü	-0.5 ile 2.5
Yanal sapma cezası	0 ile 0.025
Dikey sapma cezası	0 ile 0.05
Güç cezası	-1000 ile 4000
Kol-bacak zıt faz salınımı ödülü/cezası	-0.5 ile 1.5
Kol salınımının simetrisi cezası	0 ile 0.8

Çizelge 3.1'deki değerler, tüm tohumlar (seed) boyunca toplanmış simülasyon gözlemlerine dayanarak, eğitim sonunda elde edilmiştir; aralıkların raporlandığı yerlerde, yakınsama sonrasında tohumlar arasında gözlenen minimum-maksimum değerleri ifade eder. Ağırlıklar belirlenirken hem her bir ödül teriminin önemi hem de her bir ödül bileşeninin sayısal değer aralığı dikkate alınmış ve katsayılar bu ölçeklere göre normalize edilerek seçilmiştir. Bu normalizasyonun amacı, tüm bileşenlerin toplam ödül üzerindeki etkisini dengelemektir. Buna göre, her bir terime doğal aralığındaki maksimum değere referansla uygun bir ağırlık atandığında, hiçbir bileşen diğerlerini baskın biçimde gölgelemez ve ödül fonksiyonu çok amaçlı hedeflerini dengeli bir biçimde yerine getirir. Kazançlarıyla çarpılmadan önce, ödül fonksiyonu terimlerinin değer aralıkları Çizelge 3.1'de listelenmiştir.

Çizelge 3.2. Ödül ve ceza terimleri için ağırlık değerleri

Ödül ağırlığı	Değer
Süre ödülünün ağırlığı ($\mathrm{w_d}$)	1
İleri hız ödülünün ağırlığı ($\mathrm{w_f}$)	2
Yanal sapma cezasının ağırlığı ($\mathrm{w_l}$)	-5
Dikey sapma cezasının ağırlığı ($\mathrm{w_v}$)	-10
Güç cezasının ağırlığı (w_P)	-0.0005
Kol-bacak zıt faz salınımı ödülü/cezasının ağırlığı ($\mathrm{w_{a1}}$)	1
Kol salınımının simetrisi cezası ağırlığı ($\mathrm{w_{a2}}$)	-1

Çizelge 3.2, tüm ajanlar için aynı değerlerde kullanılan ödül ağırlıklarını listeler. Süre ödülü, robotun yürüyüşü sürdürdüğü her adımda küçük bir pozitif katkı ekleyerek bölümlerin mümkün olduğunca uzun sürmesini teşvik eder. İleri ilerleme ödülü, ileri hızı ödüllendirerek görevin temel hedefi olan y yönündeki hareketi destekler; tablodan görüldüğü üzere bu terim negatif değer alabilir, dolayısıyla geriye doğru kayma hafif biçimde cezalandırılır. Yanal ve düşey ceza terimleri, dengeyi ve yaklaşık doğrusal ilerlemeyi korumayı hedefler. Yanal sapma cezası gereksiz sağa-sola hareketi sınırlandırarak ajanın hedefe daha hızlı ulaşmasına yardımcı olurken, düşey salınım cezası zıplamayı veya aşırı yukarı-aşağı gövde hareketini bastırır. Güç cezası, tüm aktif eklemler boyunca anlık tork kullanımını cezalandırarak enerji verimliliğini zorlar. Son iki terim (kol-bacak eşzamanlı olmama ve kol simetrisi ödülü/cezası), doğal insan yürüyüşüne benzeyen kol hareketlerini teşvik eder (Heess vd. 2017; Peng vd. 2017; Radosavovic vd. 2023).

Zıt-faz koordinasyon terimi, zıt taraftaki kol ve bacak hareketlerinin birbirine zıt yönlerde olmasını ödüllendirir. Bu terim, Denklem (3.60)'da gösterildiği gibi matematiksel olarak hesaplanır:

$$r_{a1} = \theta_{Rs}\theta_{Lh} + \theta_{Ls}\theta_{Rh} \tag{3.60}$$

Burada sağ kol açısı θ_{Rs} ile sol kalça açısı θ_{Lh} ve sol kol açısı θ_{Ls} ile sağ kalça açısı θ_{Rh} gösterilmiştir. Eşitlikteki terim, normal (kontralateral, zıt-faz) kol-bacak koordinasyonunu teşvik edecek biçimde kurgulanmıştır: kollar, kontralateral kalçalarla zıt fazda salındığında, toplam büyüklük pozitif değerlere eğilim gösterir ve böylece ödül artar. Anti-normal (ipsilateral, aynı-faz) koşulu için faz tanımı, bağlanmış sinyallerin bir tarafına değerlendirme öncesinde bir işaret konvansiyonu uygulanarak doğrudan ödül-sinyali bloğu içinde gerçekleştirilir. Özellikle faz referansı, değerlendirme öncesinde kalça-açısı sinyallerinin işaretinin ters çevrilmesiyle uygulanır (yani kol-kalça çapraz bağlaşımında $-\theta_{Lh}$ ve $-\theta_{Rh}$ terimleri kullanılır). Bu işlem bir işaret-referansı (faz) konvansiyonuna karşılık gelir ve ek bir ceza getirmez veya altta yatan amaç fonksiyonunu değiştirmez. Uygulamada, kalça-açısı sinyalleri önce uzuvlar arasında ileri salınım yönü konvansiyonunu hizalamak için işaret düzeltmesine tabi tutulur (-1 ile çarpılır); bu ön işleme adımı tutarlı biçimde uygulanır ve amacı değiştirmez, yalnızca açı sinyallerinin işaret referansını değiştirir. Bu konvansiyon ile eşitlik, pozitif katkısının hedeflenen anti-normal moduna karşılık geldiği tutarlı bir koordinasyon göstergesi olarak kalır; aynı-faz davranışa karşı içsel bir ceza olarak yorumlanmaz.

İkinci kol ödül terimi olan kol-salınım simetrisi cezası, iki kol hareketinin simetrisini Denklem (3.61)'ta verildiği gibi değerlendirir:

$$r_{a2} = (\theta_{Rs} + \theta_{Ls})^2 \tag{3.61}$$

Kollar aynı yönde hareket ettiğinde, toplamın değeri büyür ve yüksek bir ceza üretilir. Buna karşılık, kollar zıt yönlerde simetrik biçimde salındığında iki kol açısının toplamı sıfıra yaklaşır ve bu terim minimum değerini alır. Yani faz farkı 180° olduğunda ceza en düşüktür. Böylece robotun kolları yürüyüş sırasında doğal bir karşı salınım sergiler; bir kol ileri giderken diğeri geri hareket eder (Collins vd. 2009; Peng vd. 2017).

Çizelge 3.3, her bir kol-hareketi koşulu için ödül formülasyonlarını özetlemektedir. Bu tablo, ortak yürüme hedeflerinin ve kol-bacak terimlerinin normal salınımı nasıl zorladığını, anti-normal (aynı-faz) modun Denklem (3.60) değerlendirilmeden önce kalça-açısı sinyalleri θ_{Lh} ve θ_{Rh}'nin işaretlerinin ters çevrilmesiyle nasıl uygulandığını (böylece aynı göstergenin hedeflenen mod için pozitif kalmaya devam ettiğini) ve sabit kol durumunda kol terimi katkılarının nasıl ihmal edilebilir hale getirildiğini göstermektedir. Kol-salınım koşulları arasında karşılaştırmanın adilliği bağlamında, buna göre ödül/Q_0 karşılaştırmaları yalnızca normal ile anti-normal arasında raporlanmış; sabit kol durumu ise kola ilişkin şekillendirme (shaping) terimlerinin çıkarılması nedeniyle koşullar arası ödül/Q_0 karşılaştırmalarına dahil edilmemiştir. Çizelge 3.3, her koşulda kullanılan kesin ödül tanımını listelemektedir. Temel yürüme hedefi tüm ayarlarda aynıdır ve ileri ilerleme, dik duruş kararlılığı ve efor düzenlemesini hedefler. Koşula bağlı tek kısım, kola ilişkin şekillendirmedir: normal ve anti-normal için amaçlanan koordinasyon modunu belirtmek üzere dahil edilir; sabit kol durumunda ise omuzlar kilitli ve tork uygulanmadığından (tahriksiz) bu terimler çıkarılır. Bu nedenle toplam getiri (total return) ve Q_0 gibi ödül tabanlı nicelikler koşul içinde yorumlanır; koşullar arası sonuçlar ise tüm politikalar için aynı biçimde hesaplanan, ödülden bağımsız eğitim sonrası (post-hoc) kararlı-durum yürüyüş ve enerji ölçütlerine dayandırılır. Özellikle, yalnızca benzetimde (simülasyonda) gözlenen hareket ve tahrik büyüklüklerine bağlı olan V_y, $|V_x|$, dz, |X|, |P|, |τ|, W_{step}, W_{dist}, ve CoT metrikleri raporlanmaktadır.

Çizelge 3.3. Üç kol-salınım koşulu için ödül fonksiyonu tanımları

Koşul	Ödül Fonksiyonu
Normal	$r^n = 2\dot{y} - 5x^2 - 10\dot{z}^2 + 0.025 - 0.0005\,P + \theta_{Rs}\theta_{Lh} + \theta_{Ls}\theta_{Rh} + (\theta_{Rs} + \theta_{Ls})^2$
Anti-normal	$r^{an} = 2\dot{y} - 5x^2 - 10\dot{z}^2 + 0.025 - 0.0005\,P + \theta_{Rs}(-\theta_{Lh}) + \theta_{Ls}(-\theta_{Rh}) + (\theta_{Rs} + \theta_{Ls})^2$
Sabit	$r^s = 2\dot{y} - 5x^2 - 10\dot{z}^2 + 0.025 - 0.0005\,P$

Not: Anti-normal kol-kalça çapraz terimindeki eksi işareti, aynı-faz koordinasyon konvansiyonunu sağlamak amacıyla kalça-açısı sinyallerine (θ_{Lh} ve θ_{Rh}) uygulanan işaret terslemesinden kaynaklanmaktadır; ek bir ceza terimi değildir ve farklı bir amaç fonksiyonu formülasyonu anlamına gelmemektedir.

Çizelge 3.3'te sunulan ödül fonksiyonlarının temel bileşenleri (ileri hız ödülü, yanal ve düşey sapma cezası, hayatta kalma sabiti ve güç cezası), Heess vd. (2017) tarafından zengin ortamlarda hareket davranışlarının ortaya çıkarılmasına yönelik önerilen ödül yapısından uyarlanmış olup ağırlık değerleri mevcut modele göre yeniden belirlenmiştir. Kol salınım koordinasyonuna ilişkin terimler ise Ayyıldız ve Polat (2026) tarafından tanımlanmıştır. Literatürde kol salınımı farklı stratejilerle ele alınmaktadır. Bazı çalışmalarda enerji minimizasyonu yoluyla açık bir kol ödülü kullanılmadan salınımın kendiliğinden oluşması hedeflenirken (Radosavovic vd., 2024), diğer çalışmalarda merkezi açısal momentum (Centroidal Angular Momentum, CAM) takibi ile dolaylı yönlendirme tercih edilmekte (Lee vd., 2025), taklit öğrenme tabanlı yaklaşımlarda ise referans hareket verisi izlenmektedir (Peng vd., 2022). Bu çalışmada benimsenen yöntem, omuz ve kalça eklem açılarının çapraz çarpımına dayanan bir koordinasyon terimi aracılığıyla istenen kol-bacak eşleşme modunu (kontralateral veya ipsilateral) doğrudan ödül fonksiyonuna yansıtmaktadır. Bu formülasyon, dinamik büyüklükler ya da referans yörünge gerektirmeden yalnızca anlık kinematik bilgiyle çalışması bakımından yukarıda sıralanan yaklaşımlardan ayrışmaktadır. Üç kol koşulunun (normal, anti-normal, sabit) biyomekanik dayanağı, Collins vd. (2009) tarafından gerçekleştirilen ve farklı kol salınım modlarının metabolik maliyet ile yer tepki momentleri üzerindeki etkisini deneysel olarak ortaya koyan çalışmaya dayanmaktadır.

3.5. Yumuşak Aktör-Eleştirmen Hiperparametreleri

İnsansı bir robotun denetimi, yüksek boyutlu eylem uzayı ve karmaşık dinamikleri nedeniyle zorludur. Daha kararlı ve daha verimli bir öğrenme elde edebilmek için, ajanın bazı hiperparametre değerlerinin çalışma bağlamına göre ayarlanması gerekir. Çizelge 3.4'te sunulan hiperparametre değerleri, tüm kol salınımı konfigürasyonları ve tüm ajan tipleri için aynı biçimde uygulanmıştır. Çizelgede ayrıca belirtilmeyen diğer parametreler ise varsayılan değerlerinde bırakılmıştır.

Çizelge 3.4. Ajan hiperparametre değerleri

Hiperparametreler	Değer
Örnekleme zamanı (denetim frekansı)	0.025
Yığın (batch) boyutu	256
Tekrar oynatma belleği uzunluğu	1e+6
Hedef yumuşatma katsayısı (Polyak katsayısı)	5e-3
Isınma adımları (öğrenme başlamadan önceki başlangıç rastgele adımları)	1024
Güncelleme başına yineleme (epoch) sayısı	2
Aktör öğrenme oranı	1e-4
Eleştirmen öğrenme oranı	5e-4
Gradyan kırpma eşiği	1
L2 düzenlileştirme katsayısı (ağırlık çürümesi)	1e-5

Kısa bir örnekleme periyodu olan 0.025 s, hızlı geri besleme sağlar ve denetim frekansını belirgin biçimde artırır (Issa ve Aldair 2022). Orta büyüklükte bir yığın (batch) kullanımı, gradyan kestirimlerini daha kararlı hâle getirir ve güncelleme başına hesaplama maliyetini düşürerek daha verimli öğrenmeye olanak tanır (Haarnoja 2018a; Hong vd. 2024). Büyük bir tekrar oynatma belleği, ajanın daha çeşitli bir durum geçmişinden yararlanmasını sağlar ve yakın deneyimlere aşırı uyum (overfitting) riskini azaltmaya yardımcı olur (Sutton vd. 1999). Seçilen hedef yumuşatma katsayısı (target smoothing factor), hedef ağların biraz daha hızlı güncellenmesini sağlar; ancak bu katsayının çok yüksek seçilmesi kararsızlığa yol açabilir (Haarnoja vd. 2018a). Başlangıç ısınma adımları (warm start), herhangi bir ağ güncellemesi yapılmadan önce toplanan rastgele etkileşim adımlarının sayısını ifade eder. Başlangıç ısınma adımlarının mini yığın (mini-batch) boyutundan büyük seçilmesi, öğrenmenin ancak tekrar oynatma belleğinde en az bir tam yığınlık deneyim biriktikten sonra başlamasını garanti eder ve bu da eğitimin erken safhalarındaki kararlılığı artırır (Louette vd. 2024). Güncelleme başına yineleme sayısının (epoch sayısı) 2 seçilmesi, her güncelleme safhasında ajanın örneklenen veri üzerinden bir kez yerine iki kez geçiş yapması anlamına gelir. Aktör ve eleştirmen için daha düşük öğrenme oranları, güncellemeleri yumuşatarak aşırı agresif adımlardan kaynaklanabilecek politika salınımlarını azaltmak amacıyla tercih edilmiştir. Seçilen gradyan eşiği (gradient threshold) ile, gradyan normu bir değerini aştığında bir değerine yeniden ölçeklendirilir; böylece patlayan gradyanlar engellenir. L2 düzenlileştirme (L2 regularization), büyük ağırlıkları cezalandırarak aşırı uyumu azaltır; büyük bir tekrar oynatma belleği ve yığın boyutu ile birlikte kullanıldığında, verinin küçük bir alt kümesinin ezberlenmesi riskini düşük tutar (Liu vd. 2011; Masters ve Luschi 2018; Tarasov vd. 2024). Her algoritma, stokastikliği kontrol etmek amacıyla beş bağımsız rastgele tohum (seed) ile eğitilmiştir (Henderson vd. 2018). Rastgele tohum; başlangıç ağ ağırlıklarını, deneyim tekrarında mini yığın örnekleme sırasını, keşif gürültüsünü ve ortamdaki olası stokastik yeniden başlatmaları belirler (Islam vd. 2017). Eğitim bütçesi ve hiperparametreler tüm koşular boyunca sabit tutulmuştur.

Bayes optimizasyonu kullanılarak, ajanın öğrenme performansını en güçlü biçimde etkileyen üç hiperparametre (mini-batch boyutu, aktör öğrenme oranı ve eleştirmen öğrenme oranı) değeri ayarlanmıştır. Algoritma farklı hiperparametre birleşimlerini dener ve ortalama bölüm ödülünü maksimize eden konfigürasyonu arar. Arama uzayı içinde bulunan en iyi değerler Çizelge 3.4'te raporlanmıştır. Birkaç rastgele değerlendirmeden sonra, Gauss süreç (Gaussian process, GP) modeli mevcut gözlemlerle güncellenerek amaç fonksiyonunun sonrasal (posterior) dağılımı elde edilir. Bu sonrasal dağılımdan, her adayın beklenen iyileşmesini ölçen bir kazanım (acquisition) fonksiyonu $\mathcal{A}(\lambda)$ tanımlanır ve bir sonraki değerlendirme noktası λ^{+} seçilmek üzere maksimize edilir. Burada $F(\lambda)$, λ hiperparametreleri ile elde edilen ortalama bölüm ödülünü; mevcut en iyi değeri ise $m^{\star}$ göstersin. Bu durumda, beklenen iyileşme Denklem (3.62)'de verildiği gibi elde edilir (Snoek vd. 2012; Gelbart vd. 2014; Shahriari vd. 2016):

$$(\lambda) = EI(\lambda) = E_{\mathrm{G}}[\max(0,\, F(\lambda) - m^{\star})]$$
$$m^{\star} = \max_{j \le t} F\left(\lambda_j\right) \tag{3.62}$$

Burada $E_G[\cdot]$, güncel GP modeli altında alınan beklenen değeri (expectation) ifade eder. GP'nin λ noktasındaki ardıl (posterior) kestiriminin $F(\lambda) \sim \mathcal{N}\left(\hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)\right)$ olduğu varsayılsın; bu durumda $EI(\lambda)$, hem tahmin edilen ortalamanın m^* değerini aşma olasılığına hem de bu aşımın büyüklüğüne dayalı bir beklenen kazanç verir. Bu nicelik, GP kestiriminden analitik olarak hesaplanabilir ve her aday λ için elde edilebilir. Her yinelemede algoritma $EI(\lambda)$'yi maksimize eden noktayı seçer, gerçek değerlendirmeyi çalıştırır, gözlemi GP modeline ekleyerek ardıl dağılımı günceller ve bir sonraki yinelemeye geçer. Süreç, yineleme sayısı veya zaman sınırına ulaşılıncaya kadar devam eder ve değerlendirmeler arasında gözlenen en yüksek ödülü sağlayan hiperparametre kümesi optimal kabul edilir.

Bu çalışmada öğrenme eğrileri ve yürüyüş zaman serileri, koşul başına beş tohum (n = 5) üzerinden, medyan ve çeyrekler arası aralık (interquartile range, IQR) kullanılarak özetlenmiştir; merkez çizgi medyanı, gölgeli bant ise IQR'yi göstermektedir. Koşullar arası farklılıklar, tohum düzeyinde özet değerler üzerinde Kruskal-Wallis (KW) testi ile sınanmıştır. KW için etki büyüklüğü η_H^2 olarak raporlanmıştır. Tablolarda verilen kararlı-durum (steady-state) metrikleri her tohum için hesaplanmış ve ardından grup düzeyinde birleştirilmiştir (grup ortalaması ve yayılım ile raporlanmıştır). Enerji/efor yorumları için, uygun durumlarda $|P|/V_y$ ve $|\tau|/V_y$ gibi normalize ölçütler kullanılmıştır (Mann ve Whitney 1947; Agarwal vd. 2022).

3.6. SAC, TD3 ve DDPG Algoritmalarının Karşılaştırılması

Bu çalışmada önerilen ES-SAC yaklaşımında, ES katmanı ile birlikte çalışan temel pekiştirmeli öğrenme algoritması SAC'tır. Ancak ES çerçevesinin algoritma-bağımsız yapısı, iç RL ajanının farklı algoritmalarla değiştirilerek performans etkisinin değerlendirilmesine olanak tanımaktadır. Bu sayede gözlemlenen başarımın yalnızca ES katmanından mı, yoksa iç RL algoritmasının kendine özgü özelliklerinden mi kaynaklandığı ayrıştırılabilir. Bu amaçla, aynı ES yapısı sabit tutularak SAC'ın yanı sıra TD3 ve DDPG algoritmaları da iç ajan olarak denenmiş ve sonuçları karşılaştırılmıştır. Bu başlık altında, ES kısmı sabit tutularak yalnızca politika yapısı, keşif stratejisi, hedef değer üretimi ve kayıp fonksiyonu güncelleme kuralları bakımından SAC, TD3 ve DDPG tabanlı ajanlar arasındaki temel farklar ele alınmıştır.

Buna karşılık TD3 ve DDPG algoritmalarında politika deterministiktir ve aktör ağı doğrudan $a = \mu_\theta(s)$ biçiminde tek bir eylem üretmektedir. Bu nedenle her iki yöntemde keşif, öğrenme sırasında deterministik eyleme harici bir gürültü terimi eklenmesiyle sağlanmaktadır. Bu süreç Denklem (3.63)'deki gibi ifade edilebilir:

$$a_t = \mu_\theta(s_t) + \varepsilon_t \tag{3.63}$$

Burada ε_t, genellikle sıfır ortalamalı Gauss veya Ornstein-Uhlenbeck tipinde seçilen keşif gürültüsünü göstermektedir. TD3 ve DDPG'de entropi terimi bulunmadığından, keşif-sömürü dengesi doğrudan bu gürültünün varyansı, korelasyonu ve kırpılma özellikleri gibi dışsal parametrelere bağlıdır.

Üç algoritmanın ortak yapısında eleştirmen ağları yer almakla birlikte, eleştirmen mimarileri ve hedef değer tanımları farklıdır. SAC ve TD3 algoritmalarında ikiz eleştirmen yapısı kullanılmakta, yani her iki yöntemde de iki bağımsız Q-ağı üzerinden değer tahmini yapılmaktadır. Buna karşılık DDPG'de yalnızca tek bir Q-ağı bulunmaktadır. Bu fark, özellikle aşırı tahmin yanlılığına karşı alınan önlemler açısından önemlidir. SAC'ta ikiz eleştirmen yapısı entropi düzenlemesiyle birlikte çalışırken, TD3'te ikiz eleştirmen yapısına ek olarak hedef politika yumuşatma uygulanmaktadır. DDPG'de ise tek eleştirmen kullanıldığından, aşırı tahmin yanlılığını azaltmaya yönelik belirgin bir özel mekanizma bulunmamaktadır. Çizelge 3.5'te "sonraki eylem seçimi" satırı incelendiğinde, SAC'ta hedef değerin hesaplanmasında kullanılan sonraki eylemin, hedef durum s_i' için mevcut stokastik politikadan örneklendiği görülmektedir. TD3'te ise hedef eylem, hedef aktör ağının çıktısına kırpılmış Gauss gürültüsü eklenmesiyle oluşturulmaktadır. Bu yaklaşım, hedef politika yumuşatma (target policy smoothing) olarak bilinmekte ve değer fonksiyonu üzerinde ani ve aşırı keskin değişimlerin etkisini azaltmaktadır. DDPG'de ise sonraki eylem yalnızca hedef aktör ağının deterministik çıktısı olarak alınmaktadır; ek bir hedef gürültü ekleme mekanizması bulunmamaktadır (Fujimoto vd. 2018; Haarnoja, Zhou, vd. 2018; Lillicrap vd. 2015).

Hedef değer üretimi bakımından da yöntemler arasında önemli farklar vardır. SAC'ta hedef değer, anlık ödüle ek olarak iki hedef eleştirmen tahmininin minimumu ve entropi terimini içeren bir biçimde oluşturulmaktadır. Bu nedenle SAC hedefi yalnızca beklenen gelecekteki getiriyi değil, aynı zamanda politikanın rastlantısallığını da hesaba katar. TD3'te hedef değer yine iki hedef eleştirmenin minimumu kullanılarak oluşturulmakta, ancak burada entropi terimi yer almamaktadır. DDPG'de ise hedef değer, tek hedef eleştirmen ağı ve hedef aktör çıktısı kullanılarak hesaplanmaktadır. Dolayısıyla SAC ve TD3, minimum operatörü sayesinde daha korumalı ve temkinli değer tahminleri üretirken, DDPG daha yalın fakat aşırı tahmine daha açık bir hedef değer yapısına sahiptir (Fujimoto vd. 2018; Haarnoja, Zhou, vd. 2018; Lillicrap vd. 2015).

Eleştirmen kaybı bakımından üç yöntemde de temel yapı, hedef değer ile ağın tahmin ettiği Q-değeri arasındaki karesel hatanın minimize edilmesine dayanmaktadır. Ancak SAC ve TD3'te bu kayıp her bir eleştirmen için ayrı ayrı tanımlanırken, DDPG'de yalnızca tek eleştirmen için tanımlanmaktadır. Bu durum, TD3 ve SAC'ın değer tahmininde daha fazla denetim sağlamasına yardımcı olmaktadır.

Aktör amacı açısından da belirgin ayrımlar mevcuttur. SAC'ta aktör amacı, entropi düzenlemeli bir optimizasyon problemidir; yani politika hem yüksek değerli eylemleri seçmeye hem de yeterli düzeyde rastlantısallığı korumaya yönlendirilir. TD3'te aktör amacı, birinci eleştirmen tarafından tahmin edilen Q-değerini maksimize edecek biçimde tanımlanır. DDPG'de de benzer biçimde aktör güncellemesi doğrudan Q-değerinin artırılmasına dayanır. Ancak TD3'ün önemli bir farkı, aktör güncellemesinin gecikmeli (delayed) biçimde yapılmasıdır. Başka bir ifadeyle, eleştirmen ağları birkaç adım güncellendikten sonra aktör ağı güncellenir. Bu sayede politika, daha olgunlaşmış değer tahminlerine göre optimize edilir. DDPG'de ise böyle bir gecikmeli güncelleme stratejisi bulunmamaktadır (Fujimoto vd. 2018; Haarnoja, Zhou, vd. 2018; Lillicrap vd. 2015).

Çizelge 3.5. SAC, TD3 ve DDPG algoritmalarının ve bunların ES tabanlı hibrit sürümleri olan ES-SAC, ES-TD3 ve ES-DDPG'nin politika yapısı, keşif mekanizması, hedef değer ve kayıp fonksiyonları açısından karşılaştırılması

Özellik/İfade	SAC	TD3	DDPG
Politika tipi	Stokastik politika: $\pi_\theta(a \mid s) = \mathcal{N}(\mu_\theta(s), \sigma_\theta(s))$	Deterministik politika: $a = \mu_\theta(s)$	Deterministik politika: $a = \mu_\theta(s)$
Keşif mekanizması	Keşif, politika dağılımının entropisi üzerinden $\sigma_\theta(s)$ ve α ile sağlanır.	Eyleme eklenen Gauss gürültüsü ε_t ile sağlanır.	Eyleme eklenen OU veya Gauss gürültüsü ε_t ile sağlanır.
Eleştirmen yapısı	İkiz Q-ağı: Q_1, Q_2	İkiz Q-ağı: Q_1, Q_2	Tek Q-ağı: Q
Sonraki eylem seçimi	$a_i' \sim \pi_\theta(\cdot \mid s_i')$	$\widetilde{a_i'} = \mu_{\bar{\theta}}(s_i') + \text{clip}(\widetilde{\varepsilon_i}, -c, c)$	$a_i' = \mu_{\bar{\theta}}(s_i')$
Hedef değer y_i	$y_i = r_i + \gamma\left(min_{k\in\{1,2\}}\overline{Q_k}(s_i', a_i') - \alpha \ln \pi_\theta(a_i' \mid s_i')\right)$	$y_i = r_i + \gamma \min 1_{k\in\{1,2\}}\overline{Q_k}(s_i', \widetilde{a_i'})$	$y_i = r_i + \gamma\bar{Q}\left(s_i', \mu_{\bar{\theta}}(s_i')\right)$
Eleştirmen kaybı	$\frac{1}{2M}\sum_i \left(y_i^{\text{SAC}} - Q_k(s_i, a_i)\right)^2$	$\frac{1}{2M}\sum_i \left(y_i^{\text{TD3}} - Q_k(s_i, a_i)\right)^2$	$\frac{1}{2M}\sum_i \left(y_i^{\text{DDPG}} - Q(s_i, a_i)\right)^2$
Aktör amacı	$\frac{1}{M}\sum_i \left[\alpha \ln \pi_\theta(a_i \mid s_i) - \min_k Q_k(s_i, a_i)\right]$	$-\frac{1}{M}\sum_i Q_1(s_i, \mu_\theta(s_i))$	$-\frac{1}{M}\sum_i Q(s_i, \mu_\theta(s_i))$
Entropi düzenlemesi	Var; α otomatik ayarlanabilir, hedef entropi $\mathcal{H}_{\text{target}}$ kullanılır.	Yok.	Yok.
Hedef politika yumuşatma	Yok; keşif doğrudan politika dağılımı üzerinden sağlanır.	Var; hedef eyleme kırpılmış Gauss gürültüsü eklenir.	Yok.
Aşırı tahmin yanlılığına karşı mekanizma	İkiz eleştirmen ve entropi düzenlemesi.	İkiz eleştirmen ve hedef politika yumuşatma.	Belirgin bir özel mekanizma yok; tek eleştirmen yapısı kullanılır.

Çizelge 3.5'in "entropi düzenlemesi" satırı da SAC ile diğer iki yöntem arasındaki temel ayrımı açık biçimde göstermektedir. SAC'ta entropi düzenlemesi doğrudan algoritmanın çekirdek parçasıdır ve α katsayısı otomatik olarak ayarlanabilir. TD3 ve DDPG'de ise bu tür bir düzenleme bulunmaz. Bu nedenle SAC, keşif davranışını içsel ve duruma bağlı bir biçimde yönetirken, TD3 ve DDPG'de keşif daha çok dışsal gürültü ekleme sürecine bağlıdır (Shen 2024).

Hedef politika yumuşatma mekanizması açısından TD3, diğer ajanlardan belirgin biçimde ayrılmaktadır. SAC'ta keşif doğrudan politika dağılımı üzerinden sağlandığı için hedef eyleme ayrıca bir gürültü eklenmez. DDPG'de de bu tür bir mekanizma yer almaz. Buna karşılık TD3'te hedef eyleme kırpılmış Gauss gürültüsü eklenerek, eleştirmenin dar bölgelerdeki hatalı ve aşırı iyimser tahminlere fazla duyarlı olması engellenmeye çalışılır. Bu özellik, TD3'ün hem hedef değer üretiminde hem de aşırı tahmin yanlılığını azaltmada önemli avantajlarından biridir (Fujimoto vd. 2018; Haarnoja, Zhou, vd. 2018; Lillicrap vd. 2015).

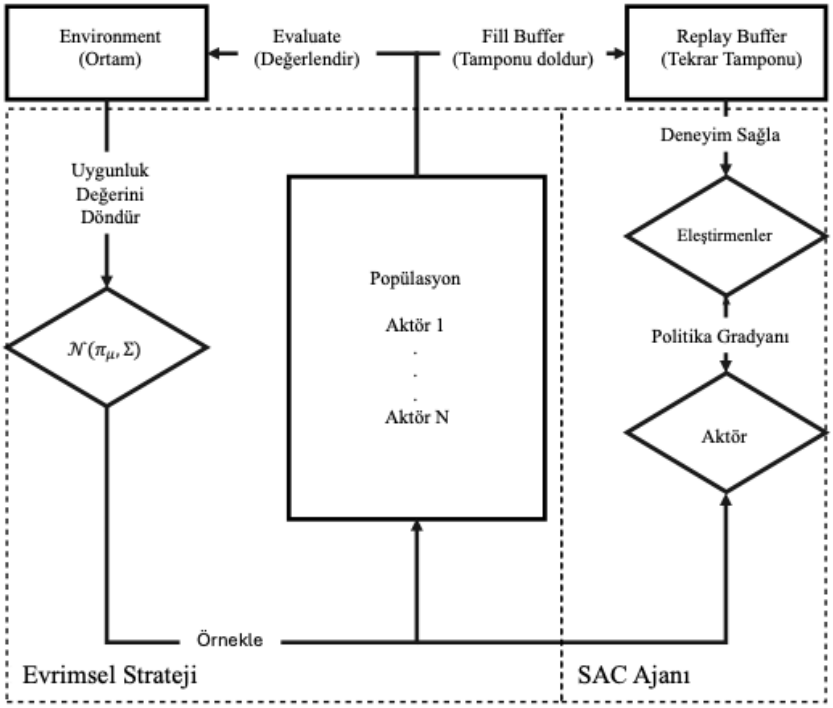
Sonuç olarak, SAC, TD3 ve DDPG algoritmaları aynı aktör-eleştirmen ailesine ait olmalarına rağmen politika temsili, keşif stratejisi, hedef değer tanımı ve güncelleme mekanizmaları bakımından belirgin biçimde ayrışmaktadır. SAC, stokastik politika ve entropi düzenlemesi sayesinde keşfi içsel olarak yöneten ve ikiz eleştirmen yapısıyla daha dengeli değer öğrenimi sunan bir yapı sergilemektedir. TD3, deterministik politika kullanmasına rağmen ikiz eleştirmen, hedef politika yumuşatma ve gecikmeli aktör güncellemesi gibi mekanizmalarla DDPG'nin zayıf yönlerini gidermeye çalışan daha sağlam bir yaklaşım sunmaktadır. DDPG ise daha sade ve daha düşük yapısal karmaşıklığa sahip olmakla birlikte, tek eleştirmenli ve entropisiz yapısı nedeniyle aşırı tahmin yanlılığına ve kararsız güncellemelere karşı daha hassastır. Bu çerçevede, Çizelge 3.5'te sunulan karşılaştırma, ES tabanlı hibrit yapılar altında SAC ve TD3 algoritmalarının DDPG'ye göre daha kararlı sonuçlar verebilmesinin olası kuramsal nedenlerini göstermektedir.

3.7. ES-SAC Hibrit Yapısının Entegrasyonu

ES-SAC algoritması, Şekil 3.9'da gösterildiği üzere, ES güçlü keşif yeteneklerini SAC yönteminin örneklem verimliliği ile birleştirerek sürekli eylem uzaylarında SAC ajanlarının daha etkin biçimde eğitilmesini sağlar. Bu yaklaşım, politika parametrelerini hem evrimsel, popülasyon tabanlı arama hem de gradyan tabanlı güncellemeler ile optimize eder, böylece ES'nin küresel arama gücünden ve SAC'nin ince ayar yapabilme kabiliyetinden eşzamanlı olarak yararlanır. Sonuç olarak ES-SAC, özellikle yüksek boyutlu durum ve eylem uzayları ile karmaşık dinamiklere sahip ortamlarda daha kararlı ve daha sağlam politikaların öğrenilmesini kolaylaştırır (Khadka ve Tumer 2018). Bu hibrit birleşim ile ES, aktörün parametre uzayında popülasyon tabanlı bir arama yürütürken SAC, yeniden oynatma arabelleğinden örneklenen deneyimler üzerinden veri uzayında gradyan temelli güncellemelerle öğrenmeyi sürdürür; böylece iki mekanizma dönüşümlü biçimde aynı eğitim döngüsünü besler.

Eğitim sürecine başlamadan önce algoritma, öncelikle ajanı ve ortamı tanımlar. Bu aşamada gözlem ve eylem uzayları belirlenir ve SAC ajanı için aktör ve eleştirmen ağları başlatılır. Daha sonra SAC ajanının aktör ağı parametreleri, ES bileşeninin entegrasyonu ile ayrıca optimize edilir. Popülasyon tabanlı bir optimizasyon yöntemi olan

ES, birden fazla ajan bireyi (worker veya individual) üretir ve bunların ortamda gösterdikleri performansı değerlendirir (Chethana vd. 2023). Hibrit ES-SAC kurgusunda ES adımı, aktör ağının parametrelerini rastgele olarak değiştirip optimize ederken, entropi katsayısı ve diğer SAC hiperparametreleri SAC safhası içinde, ondan bağımsız olarak güncellenir (Khadka ve Tumer 2018).



Şekil 3.9. ES-SAC hibrit algoritmasının genel mimarisi ve eğitim sürecindeki bileşen etkileşimleri (Pourchot ve Sigaud, 2019)

Popülasyon, evrimsel sürecin temelini oluşturur ve her bir birey politika parametrelerinin farklı bir kümesini temsil eder. Başlangıçta popülasyon, belirli bir ortalama ve standart sapma etrafında merkezlenmiş rasgele örneklenmiş bireylerden oluşur. Bu ortalama (Gaussian'ın merkezi), ilgili jenerasyonda SAC aktörünün güncel taban parametrelerine karşılık gelir; bireyler ise bu merkez etrafında örneklenen küçük parametre varyasyonlarıdır. Her bireyin performansı, ortamda çalıştırılarak elde edilen kümülatif ödül ile ölçülür. Değerlendirmenin tek bir koşuya bağlı kalması, ortamın stokastikliği ve ölçüm gürültüsü nedeniyle yanıltıcı olabileceğinden, uygunluk değeri çoğunlukla birden fazla bağımsız koşunun (farklı başlangıç koşulları veya tohumlar) ortalaması alınarak hesaplanır. Bu ortalama alma, rastlantısal dalgalanmaları bastırarak politikanın gerçek performansına daha güvenilir bir kestirim sağlar. Bir ajanın belirli bir nesilde elde ettiği toplam ödül, o birey için uygunluk değeri olarak kullanılır. Daha sonra en yüksek ödülü alan bireyler seçilerek popülasyonun bir sonraki nesli oluşturulur (Liu vd. 2024). Bu yinelemeli süreç, dağılımın merkezini (popülasyon ortalamasını) elit bireylerin bulunduğu bölgeye doğru kaydırarak aramayı yüksek uygunluk bölgelerine yönlendirmektir.

ES güncellemeleri, parametre uzayında popülasyon ortalamasını ve varyansını optimize ederek gerçekleştirilir. Başarılı bireyler, kendilerine atanan ağırlıklara bağlı olarak popülasyon ortalamasına katkıda bulunur. Yeni bireyler üretilirken, bir sonraki nesli oluşturmak üzere bu ortalamaya normal dağılımdan örneklenen rasgele gürültü eklenir. Bu gürültünün büyüklüğü, keşif ve sömürü (exploration ve exploitation) arasındaki dengeyi sağlamada kritik bir rol oynar. Gürültü seviyesinin artırılması popülasyonun çözüm uzayını daha geniş biçimde keşfetmesine olanak tanır, ancak aynı zamanda öğrenilen politikalardan daha kararsız eylemler üretilmesine de yol açabilir. Pratikte ES sürecindeki standart sapma parametreleri, popülasyon güncelleme kuralı ve ilgili hiperparametreler tarafından kontrol edilir ve aramanın geniş keşiften daha kararlı çözümlere doğru kaydırılması için zamanla kademeli olarak azaltılabilir (Li vd. 2024).

SAC ajanı, ES tarafından optimize edilen popülasyondan elde edilen en iyi politika parametrelerini kullanarak öğrenme sürecine devam eder. Bu süreçte ES safhası yalnızca aktör parametrelerine etki eder; eleştirmen ağlarının güncellenmesi SAC safhasında, yeniden oynatma arabelleğinden örneklenen verilerle gradyan temelli olarak gerçekleştirilir. Popülasyon tarafından toplanan deneyimler, SAC'nin tekrar oynatma belleğinde saklanır ve böylece ajan geçmiş etkileşimlerden öğrenebilir. Bu sayede yalnızca en iyi bireyin değil, popülasyonun tamamının ürettiği deneyim birikimi SAC güncellemelerine taşınır ve örneklem verimliliği artar. SAC ajanı eleştirmen ağları üzerinden geri bildirim alır ve entropi ile düzenlenmiş amaç fonksiyonunu maksimize ederek politika güncellemelerini gerçekleştirir, kararlılık ise hedef eleştirmen ağları aracılığıyla iyileştirilir (Khadka ve Tumer 2018; Pourchot ve Sigaud 2019).

Eğitim sırasında iki bileşen şu biçimde etkileşir: ES, parametrelerine gürültü eklenmiş aktörlerden oluşan bir popülasyonu değerlendirerek keşif alanını genişletir, SAC ise tekrar oynatma belleği üzerinden aktörü ve eleştirmenleri güncelleyerek örneklem verimliliğini ve politika optimizasyonunu artırır. En güncel SAC aktörü, belirli aralıklarla ek bir birey olarak ES popülasyonuna yeniden enjekte edilebilir, böylece evrimsel arama ve gradyan tabanlı öğrenme karşılıklı olarak birbirini besler (Fogel ve David 2006; Khadka ve Tumer 2018; Pourchot ve Sigaud 2019; Lillicrap vd. 2015; Farinelli vd. 2020; Dong ve Li 2024; Lin vd. 2024). ES adımı aktörü uygun bir bölgeye taşıdığında, SAC güncellemeleri bu çözümü ince ayarla daha da iyileştirir. Buna karşılık, pertürbasyonların geçici performans kaybına yol açtığı durumlarda eleştirmenlerin değer kestirimleri, SAC güncellemelerinin politikayı yeniden daha elverişli bölgelere yönlendirmesine yardımcı olan dengeleyici bir işlev görür.

ES-SAC hibrit sisteminde ES algoritması, RL problemini bir kara kutu optimizasyon görevi olarak ele alır ve politika parametrelerini buna göre optimize eder. Aktör ağı tarafından temsil edilen politika popülasyonu, ES adımı sırasında parametrelere gürültü eklenerek örneklenir (Hu vd. 2024). Her yavru politika (birey), kümülatif ödüllere göre hesaplanan bir uygunluk değeri elde etmek için görev ortamında değerlendirilir. ES algoritması, uygunluk değerlerine göre en iyi performans gösteren elit politikaları seçer ve bunları bir sonraki popülasyonu üretmek için kullanır (Lin vd. 2024).

Hibrit eğitim modelinde SAC, parametrelerini gradyan inişi ile ya ayrı bir zaman ölçeğinde ya da paralel olarak güncellerken, ES tarafında güncelleme süreci gradyan tabanlı değildir. Bunun yerine SAC ve ES politikaları sistem içinde birbirleriyle etkileşir. SAC tarafından en güncel biçimde eğitilen politika ES popülasyonuna dahil edilir ve ES

ortamında değerlendirilir. Bu sayede ES, farklı politika adaylarından oluşan çeşitli bir küme sunarak keşfi artırır. SAC yerel optimumlarda sıkıştığında veya yüksek boyutlu ödül yapılarını keşfetmekte zorlandığında, ES politikaları bu zorlukların aşılmasına yardımcı olabilir (Marchesini vd. 2020; Dong ve Li 2024). Bu etkileşim, yakınsamayı hızlandırır ve SAC algoritmasının arama uzayının alt optimal bölgelerinde sıkışıp kalma olasılığını azaltır. Önemli olarak, bu süreç boyunca SAC ve ES'nin temel algoritmik yapıları değişmeden korunur (Suri vd. 2020).

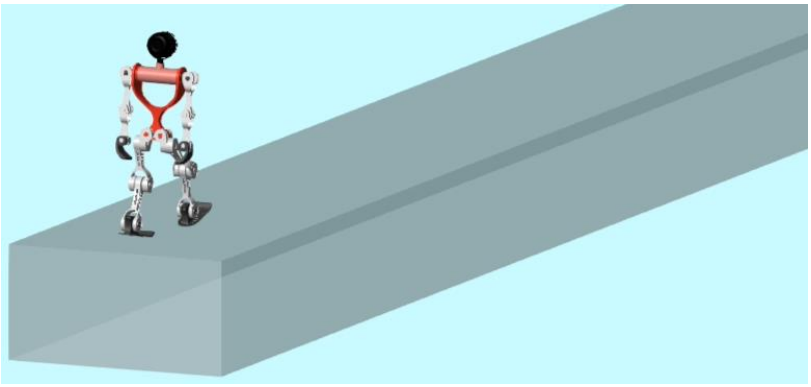
SAC yöntemi tek başına, farklı politikalardan elde edilmiş önceki deneyim verilerini tekrar kullanarak öğrenmeye devam edebilen yapısı ve entropi temelli keşif kabiliyeti sayesinde tatmin edici bir performans sergiler. Buna karşılık ES, aynı anda çok sayıda aday politikanın çalıştırılabilmesine (paralelleştirme) olanak tanıyan yapısı sayesinde yüksek boyutlu problemler için güçlü bir ölçeklenebilirlik sunar (Suri vd. 2020; Suri 2022). SAC ve ES'nin hibrit bir ES-SAC ajanında birleştirilmesi ile yaklaşım, optimal eylemlere yakın davranışlar seçmek için her iki yöntemin güçlü yönlerinden yararlanır. Çok sayıda çalışma ve bu çalışma kapsamındaki bulgular, bu tür hibrit sistemlerin, tek başına kullanılan DRL yöntemlerine kıyasla çoğu zaman daha iyi performans gösterdiğini ortaya koymaktadır(Khadka ve Tumer 2018; Suri vd. 2020; Liu vd. 2024).

4. BULGULAR VE TARTIŞMA

MATLAB/Simulink ortamında, RL Toolbox, Deep Learning Toolbox ve Simscape Multibody kullanılarak bir simülasyon sistemi geliştirilmiştir. Robotun her bir bacağında, dengeyi sağlamak ve yürüme hareketini yönlendirmek üzere tork denetimli dönel eklemler (kalça, diz ve ayak bileği) kullanılmıştır; bu yapı, insanların adım atarken kaslarını bilinçli biçimde kullanma mekanizmasını taklit etmektedir. Bu çalışmada deneysel değerlendirme iki aşamalı bir yapı içinde kurgulanmıştır.

4.1. Düz Zeminde Denetleyici Performanslarının Karşılaştırılması

İlk aşamada, kol etkisini dışlayarak yalnızca yürüme davranışının temel performansını incelemek amacıyla kollar sabitlenmiş ve robotun yürüme analizi gerçekleştirilmiştir. Bu sebepten dolayı kollar doğal salınım ve denge hissi sağlamak amacıyla sagittal ve frontal düzlemlerde pasif dönme serbestliğine sahip omuz eklemleriyle yapılandırılmıştır. Simülasyon boyunca model, temas kuvvetlerini, gövdenin uzaydaki konumunu ve yönelimini, eklem açılarını ve ileri yönlü hareketi sürekli olarak izleyerek çevre ile etkileşimlerini yorumlayabilecek şekilde çalıştırılmıştır. Eğitim düz bir yüzey üzerinde gerçekleştirilmiş ve ajanların ödülleri bu hat boyunca karşılaştırılmıştır. MATLAB ortamında Birleşik Robot Tanımlama Formatı (Unified Robot Description Format, URDF) tabanlı bir insansı robot modeli kullanılmış; eğitim işlemleri yalnızca Şekil 4.1'de gösterildiği gibi düz bir parkur üzerinde yürütülmüştür. Bu aşamada, SAC, TD3 ve DDPG ajanları ile bu yöntemlerin ES tabanlı hibrit sürümleri olan ES-SAC, ES-TD3 ve ES-DDPG ajanları karşılaştırılmıştır. Karşılaştırmalar sırasında 6 ajanın kendi aralarındaki hem pozitif hem de negatif davranışları değerlendirilmiş; kararlılık, ilerleme başarımı ve yürüyüş karakteristiği açısından yöntemler arasında ayrıntılı bir performans analizi yapılmıştır. Bu ilk aşamanın temel amacı, kol hareketinden bağımsız olarak en uygun denetleyici yapısını belirlemek ve sonraki analizler için güvenilir bir temel ajan seçmektir.



Şekil 4.1. 50 m'lik düz parkur üzerinde hareket eden insansı robot

Eğitim ve simülasyon süreçleri boyunca ileri yön (y-ekseni), yanal yön (x-ekseni) ve düşey yön (z-ekseni), robotun gövde koordinat sistemi referans alınarak

tanımlanmıştır. Ajanın ödül fonksiyonu; ileri hız için doğrudan ödül, yanal ve düşey sapmalar için (karesel) ceza terimleri ve güç tüketimi ile ayakta kalma süresini dikkate alan bileşenler içermektedir. Bu yaklaşım, dengeyi koruyan, enerji verimliliğini destekleyen ve ileri yönlü hareketi teşvik eden öğrenme davranışlarının geliştirilmesini amaçlamaktadır. Bu tür bir ödül fonksiyonu yapısı, insansı robot literatüründe yaygın olarak kullanılmaktadır (Heess vd. 2017; Radosavovic vd. 2023, 2024; Jung vd. 2025; Wu vd. 2025). Bu çalışmada kullanılan ödül fonksiyonu Heess vd. (2017) çalışmasından uyarlanmış, ancak katsayı değerleri ve bazı terimler kullanılan benzetim düzenine uygun olacak biçimde değiştirilmiştir. Bu katsayıların belirlenmesinde, her bir ödül bileşeninin göreli önemi ile sayısal değer aralıkları birlikte değerlendirilmiş ve ağırlıklar buna göre normalize edilmiştir. Katsayı değerleri başlangıçta sabitlenmiş, ajan bazında ayrı ayrı ayarlanmamış ve böylece karşılaştırma aynı amaç fonksiyonu altında öğrenme kurallarına odaklanacak biçimde kurgulanmıştır. Tüm ajanlarda kullanılan hiperparametre değerleri, ilgili başlık altında sunulan Çizelge 4.1'de verilen değerlerle aynı olacak biçimde sabit tutulmuştur. Böylece yöntemler arası karşılaştırmanın, hiperparametre farklılıklarından değil, doğrudan öğrenme algoritmalarının yapısal özelliklerinden kaynaklanan performans farklarına dayanması hedeflenmiştir. Başka bir ifadeyle, adil ve tutarlı bir karşılaştırma zemini oluşturmak amacıyla tüm ajanlar aynı deneysel koşullar altında değerlendirilmiştir. Parametre değerlerinin seçimi sırasında, yalnızca algoritmanın yakınsama başarımı değil; aynı zamanda öğrenme kararlılığı, sayısal güvenilirlik, genelleme yeteneği ve simülasyon ortamına uyum gibi ölçütler de dikkate alınmıştır. Bu kapsamda, örnekleme zamanı (sampling time), robotun çevresel değişimlere daha hassas ve zamanında tepki verebilmesini sağlar; batch size (yığın boyutu), öğrenmenin daha kararlı ve daha hızlı ilerlemesini destekleyerek politikanın düzenli biçimde iyileşmesine ve ani dalgalanmaların önlenmesine yardımcı olur; deneyim arabelleği uzunluğu (experience buffer length), politikanın genelleme kapasitesini artırır; Hedef yumuşatma katsayısı (target smooth factor), Q hedeflerinin güncel kalmasını sağlayarak özellikle dinamik ortamlarda ajan uyarlanabilirliğini iyileştirir; warm-start adımları, başlangıç öğrenme eğrisindeki ani düşüşleri ve kararsız politika oluşumunu önler ve böylece yeniden oynatma arabelleği daha güvenilir verilerle doldurulur. Güncelleme başına epoch sayısı (epochs per update), aynı veri üzerinde birden fazla güncelleme yapılmasına olanak tanır ve politikanın her mini-batch'ten mümkün olan en yüksek bilgiyi elde etmesini sağlar; düşük öğrenme oranlarının kullanılmasıyla yavaş fakat kararlı bir öğrenme süreci elde edilir; gradyan eşiği (gradient threshold), patlayan gradyan (exploding gradient) problemini azaltırken, L2 düzenlileştirme katsayısı gereksiz ağırlık büyümesini engeller ve aşırı öğrenmeyi (overfitting) azaltır. Tek bir ES jenerasyonu içinde, sıfır ortalamalı Gaussian gürültü ile 25 aday politika değerlendirilir; yani ağın ağırlık ve önyargı parametrelerine 0.25 mutasyon adım büyüklüğü uygulanır ve elit bireyler seçilir. Aynı jenerasyon içinde ajan, bu iyileştirmeleri gradyan tabanlı güncellemelerle pekiştirmek amacıyla 50 iç eğitim döngüsü (inner training cycle) çalıştırır ve böylece bir sonraki jenerasyon için rafine edilmiş bir politika elde edilir. Bu prosedür 1500 jenerasyon boyunca tekrarlanır. MATLAB ortamındaki robot simülasyonu ve eğitimi, evrimsel sürecin algoritmik yapısı nedeniyle eşzamansız (asynchronous) olarak yürütülememiştir. Bu nedenle robota yalnızca eşzamanlı (synchronous) eğitim uygulanmıştır. Tüm eğitim ve test işlemleri, yüksek performanslı bir iş istasyonunda (16 çekirdekli AMD Ryzen 9 9950X3D işlemci, NVIDIA GeForce RTX 5090 GPU, 64 GB RAM) yürütülmüştür. Her bir koşul, ilk çalışma için 1500 jenerasyon, ikinci çalışma için ise 750 jenerasyon boyunca eğitilmiştir; bölüm (episode) uzunluğu ise azami adım sayısı ile sınırlandırılmıştır.

Toplam eğitim süresini azaltmak amacıyla, eğitim sırasında bireylerin paralel değerlendirilmesi etkinleştirilmiştir.

ES parametre ayarları aşağıdaki biçimde belirlenmiştir: Elit oranı %50 olarak seçilmiştir. Buna bağlı olarak en iyi 12 birey elit olarak tanımlanmış ve bir sonraki jenerasyona aktarılmıştır. Döndürülen Politika (Returned Policy), eğitim sonunda en yüksek performansı gösteren politika olarak tanımlanmış ve nihai çözüm olarak kaydedilmiştir. Popülasyon güncellemesinde Ağırlıklı Birleştirme (Weighted Mixing) yaklaşımı kullanılmıştır. Buna göre her bir sonraki jenerasyonda yeni politikalar, elit parametrelerin ağırlıklı birleşimleri şeklinde oluşturulmuştur. Eğitim dönemi (Train Epochs) parametresi, her bir bireyin jenerasyon başına gerçekleştirdiği SAC politika güncelleme sayısını ifade etmektedir. Bu çalışmada bu değer 50 olarak ayarlanmış, böylece her bireyin bir jenerasyon içinde öğrenme için yeterli fırsata sahip olması sağlanmıştır. Başlangıç Standart Sapması (Initial Standard Deviation) 0.25 olarak belirlenmiştir. Bu seçim, parametre pertürbasyonlarında yeterli çeşitliliği korurken öğrenme sürecinde aşırı salınımların önlenmesini amaçlamaktadır. Birey Başına Değerlendirme Sayısı (Evaluations Per Individual) parametresi 1 olarak ayarlanmıştır. Yani her birey tek bir benzetim yürütmesi üzerinden değerlendirilmiştir. Bu tercih, birey başına değerlendirme süresini kısaltmış ve buna bağlı olarak toplam eğitim süresini önemli ölçüde azaltmıştır. Bölüm (Episode) uzunluğu tüm koşullar için sabit tutulmuş ve benzetim ufku ile örnekleme zamanından hesaplanmıştır.

Çizelge 4.1, hibrit ve saf pekiştirmeli öğrenme ajanlarının ortak eğitim ufku altındaki karşılaştırmalı performans sonuçlarını sunmaktadır. Sonuçlar, her bir yöntem için 5 farklı rastgele tohumdan elde edilen deneyler üzerinden hesaplanmıştır. Çizelgede yer alan Eğri Altı Alan değeri (Area Under the Curve, AUC), öğrenme eğrisi altında kalan toplam alanı, başka bir ifadeyle eğitim süreci boyunca biriken toplam ödülü göstermektedir. Sınıflandırma problemlerinde kullanılan 0-1 ölçekli AUC'dan farklıdır. Normalize Eğri Altı Alan (nAUC) ise AUC değerinin eğitim uzunluğuna bölünmesiyle elde edilmekte ve bölüm başına ortalama ödülü temsil etmektedir. Bu nedenle nAUC, özellikle aynı veya benzer eğitim ufku altındaki yöntemlerin genel öğrenme performansını karşılaştırmak için daha doğrudan bir ölçüt sunmaktadır. Son 50 Bölüm Ödülü, eğitimin son kısmındaki kararlı durum performansını; En İyi 50 Bölüm Ortalaması ise ardışık en iyi 50 bölümlük pencerenin ortalama ödülünü, yani sürdürülebilen tepe performansı göstermektedir. Hedefe Ulaşma Süresi, her yöntemin kendi en iyi 50 bölüm ortalaması referans alınarak tanımlanan %80 yeterlilik düzeyine ulaşmak için gerekli bölüm sayısını ifade etmektedir. Böylece bu ölçüt, yalnızca nihai performansı değil, aynı zamanda öğrenmenin ne kadar hızlı ve ne kadar istikrarlı biçimde yeterli düzeye ulaştığını da göstermektedir. Bu denetim hızında (T_s = 0.025 s) ve bölüm üst sınırı T_f = 60 s iken, her bölüm T_f/T_s = 2400 denetim adımından oluşmaktadır. 1500 jenerasyon boyunca bu değer, koşum başına 1500×25×2400 = 90 milyon çevre etkileşimine karşılık gelmektedir (iç gradyan güncellemeleri hariç; 38.4 milyon örnek geçişi/sample-passes hariç tutulmuştur). Genel öğrenme performansını temsil eden nAUC ve AUC ölçütlerinde en yüksek sonuç ES-SAC yöntemi tarafından elde edilmiştir. ES-TD3, ES-SAC'in gerisinde kalmakla birlikte ikinci en güçlü hibrit yöntem olarak öne çıkmaktadır. ES-DDPG ise diğer iki hibrit yaklaşıma göre belirgin biçimde daha düşük öğrenme performansı göstermektedir. Buna karşılık saf RL taban çizgilerinin nAUC ve AUC değerleri hibrit yöntemlere kıyasla oldukça düşük kalmış, dolayısıyla bu görev yapısında evrimsel arama

ile desteklenen hibrit yaklaşımın belirgin bir avantaj sağladığı görülmüştür. Kararlı durum ve tepe performansı birlikte değerlendirildiğinde, ES-SAC hem Son 50 Bölüm Ödülü hem de En İyi 50 Bölüm Ortalaması bakımından en yüksek değerlere ulaşmaktadır. ES-TD3, bu ölçütlerde ES-SAC'e görece yakın sonuçlar üretse de genel olarak daha yüksek öğrenme başarısı sergileyememiştir. Hedefe Ulaşma Süresi bakımından da ES-SAC, daha düşük ortalama değer ve daha sınırlı yayılım ile yeterlilik düzeyine en hızlı ve en tutarlı biçimde ulaşan yöntem olarak öne çıkmaktadır. ES-TD3 ise ortalama olarak hedefe daha geç ulaşmakta ve tohumlar arasında daha geniş bir değişkenlik göstermektedir. Görev tamamlama performansı açısından değerlendirildiğinde, 50 m parkurun en kısa sürede ES-TD3 ile tamamlandığı görülmektedir. ES-SAC'ın tamamlama süresi ES-TD3'e göre daha uzun olmaktadır. Diğer ajanlar ise 50 m parkuru verilen görev koşulları altında tamamlayamamıştır. Bu sonuç, ortalama ileri hız bakımından ES-TD3'ün ES-SAC'ten daha yüksek değerlere ulaşmasıyla uyumludur. Bununla birlikte, parkur tamamlama süresi ve ileri hız tek başına yeterli performans ölçütleri değildir. Bu büyüklükler, gövde dengesi, yanal sapma, tork davranışı ve denetim kararlılığı ile birlikte değerlendirilmelidir. Nitekim daha yüksek hız bazı durumlarda daha agresif denetim davranışı ile ortaya çıkabilmektedir. Bu nedenle değerlendirme yalnızca hız temelli değil, öğrenme başarısı, kararlılık ve görev güvenilirliği birlikte dikkate alınarak yapılmalıdır.

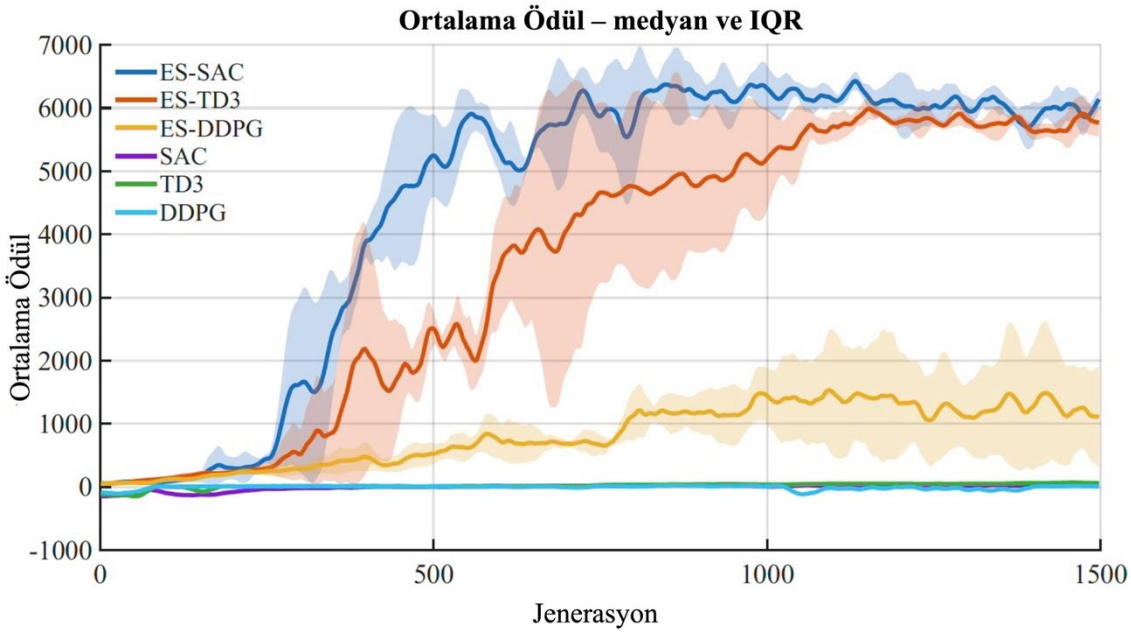
Çizelge 4.1. Hibrit ve saf RL ajanlarının karşılaştırmalı öğrenme ve görev performansları

Ajan	nAUC ($\mu \pm SD$)	AUC ($\times 10^6$)	Son 50 Bölüm Ödülü	En İyi 50 Bölüm Ortalaması	Hedefe Ulaşma Süresi	50 m Tamamlama Süresi
ES-SAC	4536,50 ±166,15	6,80 ±0,25	6015,01 ±216,93	6479,51 ±318,91	454,50 ±40,31	47.72
ES-TD3	3554,50 ±599,98	5,33 ±0,90	5825,82 ±251,86	5961,64 ±78,03	635,00 ±455,38	36.35
ES-DDPG	830,73 ±255,03	1,25 ±0,38	1179,72 ±699,98	1620,66 ±729,57	608,00 ±207,89	-
SAC	1.23 ±4.85	0.002 ±0.0073	40.02 ±16.36	43.78 ±11.82	515 ±73	-
TD3	17.81 ±4.18	0.027 ±0.0063	70.90 ±16.61	74.79 ±18.98	987 ±167	-
DDPG	-1.37 ±8.46	−0.002 ±0.013	5.93 ±27.40	29.87 ±11.76	261 ±310	-

Not: nAUC = AUC/1500. Hedefe Ulaşma Süresi, her yöntemin kendi en iyi 50 bölüm ortalamasının %80'ine ulaşmak için gereken bölüm sayısıdır. Tamamlama süresi yalnızca parkuru başarıyla tamamlayan ajanlar için verilmiştir.

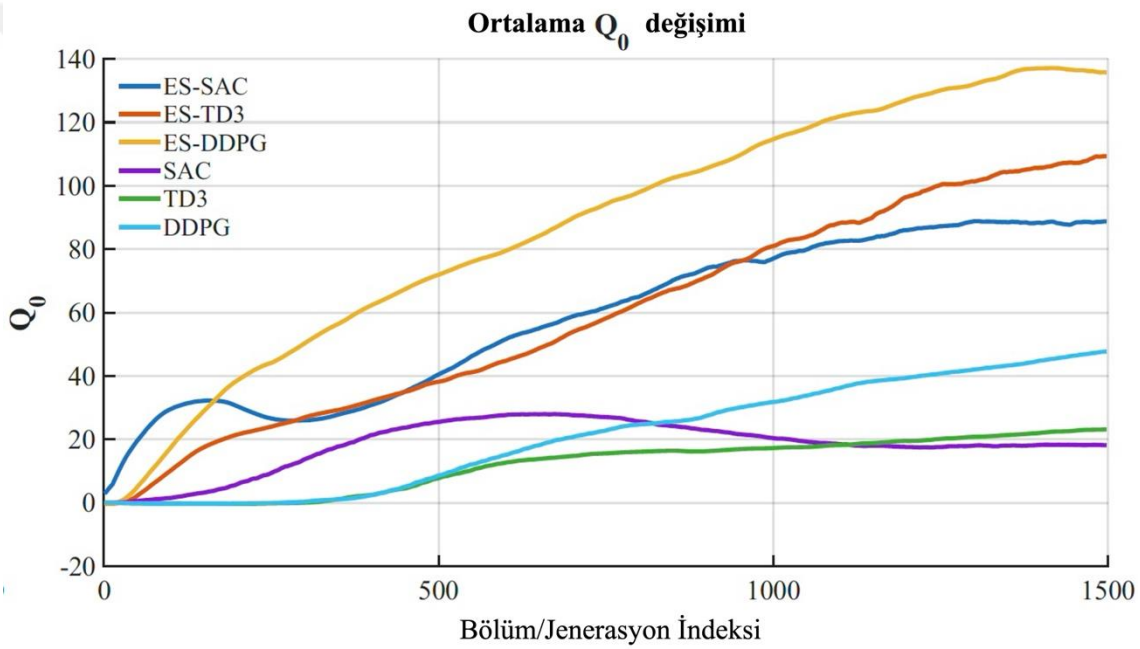
Şekil 4.2-4.5 arasında kullanılan renk şeması şu şekildedir: ES-SAC (mavi), ES-TD3 (turuncu), ES-DDPG (sarı), SAC (mor), TD3 (yeşil), DDPG (camgöbeği); gölgeli bantlar (varsa) ortalama ± standart sapmayı (standard deviation, SD) göstermektedir. Raporlanan tüm büyüklükler (örneğin yanal ve düşey sapmalar, ortalama torklar vb.) rastgele tohumlar (random seeds) boyunca ortalama ± standart sapma olarak sunulmuştur.

Şekil 4.2, eğitilen ajanlara ait öğrenme eğrilerini göstermektedir; gölgeli bantlar ortalama ± SD'yi ifade etmektedir. Eğitim, düz bir yüzey üzerinde ve robotun kolları yürüyüş sırasında serbest salınım yapacak biçimde gerçekleştirilmiştir. Şekil, ES-SAC ajanının diğerlerine kıyasla tutarlı biçimde daha yüksek ortalama ödül elde ettiğini açıkça göstermektedir. Dikkat çekici olarak, eğitim sürecinin 567. bölümünde ES-SAC ajanı 6000 ortalama ödül seviyesini aşan ilk ajan olmuş ve böylece yüksek performans "eşiğine" ulaşmıştır. Bu noktaya kadar ES-SAC ajanı diğerlerine göre daha yüksek bir öğrenme ivmesi sergilemiş, bu da algoritmanın çevresel dinamiklere hızlı ve etkili biçimde uyum sağlama kapasitesini ortaya koymuştur. ES-SAC ve ES-TD3 ajanları ilerleyen bölümlerde öğrenme doygunluğuna ulaşmış olsa da diğer ajanlar bu eşiğe ulaşamamış ve çok daha düşük ortalama ödüllerde kalmıştır. Ayrıca ES-TD3 en geniş gölgeli banda (±SD) sahiptir; bu durum tohumlar arası değişkenliğin daha fazla olduğunu ve öğrenmenin daha az kararlı seyrettiğini göstermektedir. ES-SAC daha dar bir bant göstermekte, bu da daha istikrarlı ve güvenilir bir yakınsamaya işaret etmektedir. ES-DDPG ise küçük bir SD'ye sahip olmasına rağmen düşük bir ortalama ödül üretmektedir; yani performansı tutarlı biçimde düşüktür.



Şekil 4.2. Düz zeminde serbest kol salınımı koşulunda eğitilen ajanlara ait öğrenme eğrileri (ortalama ± standart sapma)

Şekil 4.3 tüm ajanlar için Q_0 değerlerini göstermektedir. Q_0 değeri, ajanların bir bölümün başlangıcında beklenen gelecekteki ödüle ilişkin tahminini yansıtır ve öğrenme sürecine duydukları güveni gösterir. ES-DDPG ajanı oldukça iyimser Q_0 öngörüleri sergilese de aşırı tahmin eğilimi göstermiş ve düşük gerçek ödüller elde etmiştir. Buna karşılık ES-SAC ve ES-TD3 ajanları Q_0'da daha temkinli fakat istikrarlı artışlar göstermiş ve bunun sonucunda daha iyi gerçek ödüller üretmiştir. Bu durum, söz konusu ajanların öğrenme süreçlerinin daha dengeli ve güvenilir olduğunu düşündürmektedir. ES-SAC ve ES-TD3 ajanlarının ortalama Q_0 değerleri birbirine yakın olup sırasıyla 58.507 ve 59.217 olarak ölçülmüştür. Klasik (vanilla) ajanlar arasında DDPG, düşük getirilere rağmen Q_0'da orta düzeyde bir yükselme eğilimi göstermiştir; buna karşın SAC ve TD3 düşük seviyelerde görece yatay seyretmiş, bu da sınırlı öğrenme ile tutarlıdır.



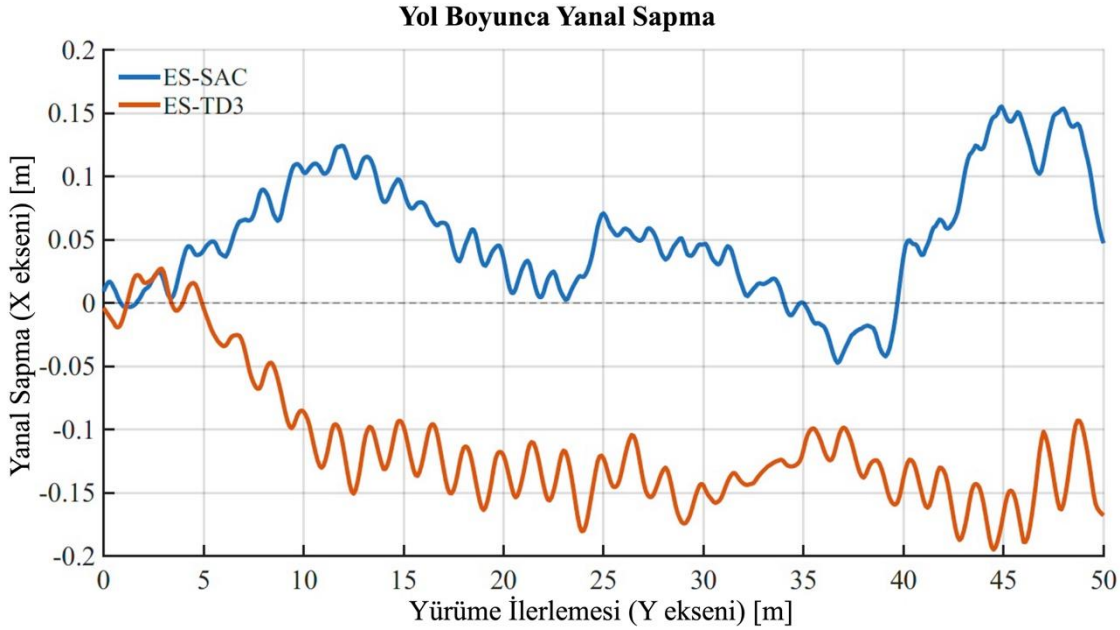
Şekil 4.3. Düz zeminde serbest kol salınımı koşulunda eğitilen ajanlara ait Q_0 değerlerinin bölüm sayısına göre değişimi

Eğitim sürecinde ajanların performansları öncelikle ortalama ödül ve Q_0 değerleri kullanılarak değerlendirilmiştir. Ancak daha derin içgörüler elde etmek amacıyla, robotun benzetim ortamındaki gerçek yürüme davranışı da incelenmiştir. Bu kapsamda özellikle eklemlerdeki ortalama tork değerleri ile gövdenin yatay ve düşey sapma miktarları ölçülmüş ve analiz edilmiştir.

Tork verimliliği robotun enerji tüketimini doğrudan etkilediğinden, denetim stratejisinin enerji verimliliğine ilişkin önemli bilgiler sunar. Benzer biçimde yatay sapma, robotun dengesini koruma başarısını ve hedef doğrultusunda ne kadar kararlı

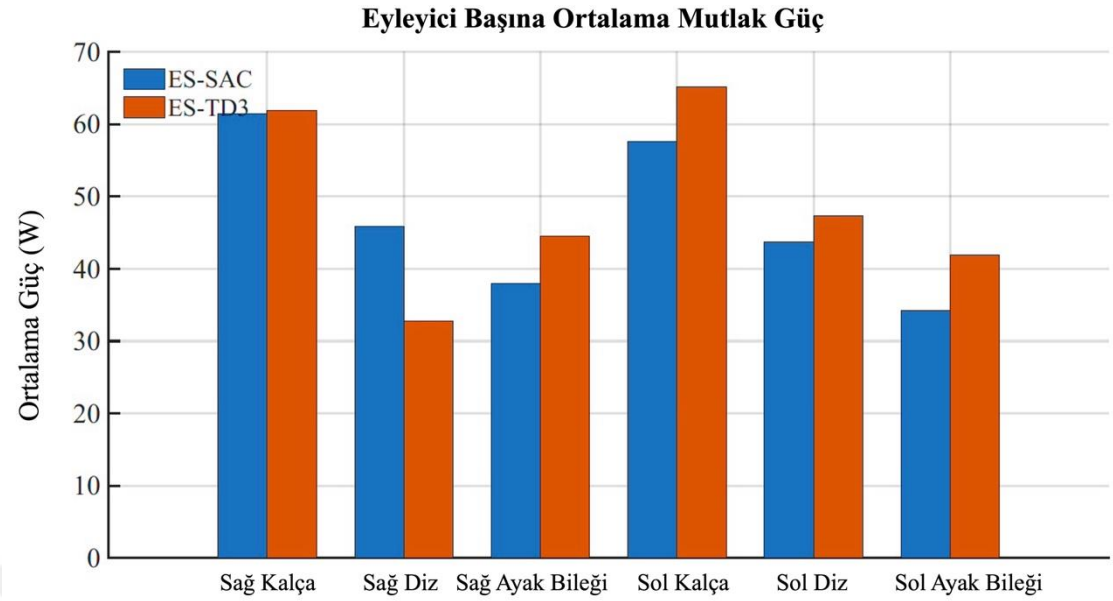
ilerlediğini değerlendirmek için kritik bir ölçüttür. Bu iki parametre birlikte ele alındığında, yalnızca elde edilen ödül değerleri hakkında değil, aynı zamanda ajanın fiziksel davranış kalitesi hakkında da önemli ipuçları sağlar. Düşey sapma analizleri, her iki ajanın da robotun ağırlık merkezini yere yakın tutarak dengeyi koruma eğiliminde olduğunu göstermektedir. Ortalama düşey sapma ES-SAC için −0.0285 m, ES-TD3 için ise −0.0158 m olarak hesaplanmıştır. Buna karşılık mutlak sapma değerleri sırasıyla 0.0674 m ve 0.1177 m olup, ES-TD3 ajanının biraz daha yüksek düşey değişkenlik sergilediğini göstermektedir. Bu veriler, her iki ajanın da düşey doğrultuda kararlı yürüyüş örüntüleri ürettiğine işaret etmektedir; ancak ES-SAC ajanı ağırlık merkezini biraz daha aşağıda tutarak dengeyi daha temkinli biçimde kontrol etmiştir.

Bu grafiklerin amacı, her bir ajanın yürüyüş sürecindeki kararlılık, verimlilik ve denetim başarısını görsel olarak ortaya koymaktır. Böylece yalnızca sayısal ödül değerlerine dayalı değil, fiziksel denge ve enerji optimizasyonu açısından da karşılaştırmalı bir değerlendirme yapılabilmektedir.



Şekil 4.4. ES-SAC ve ES-TD3 ajanları için yürüyüş yolu boyunca yanal sapmanın karşılaştırılması

Şekil 4.4 robotun yatay sapmasını göstermektedir. Daha yüksek sapma, robot için daha yüksek ceza anlamına gelmektedir. İdeal yol, gri kesikli çizgi ile gösterilmiştir. Robot gövdesinin yatay sapmaları karşılaştırıldığında, ES-SAC ajanının ideal yola daha yakın kalarak açıkça daha iyi performans gösterdiği görülmektedir. Ortalama mutlak sapma değerleri ES-SAC ajanı için 0.0576 m, ES-TD3 ajanı için ise 0.1146 m olarak bulunmuştur.



Şekil 4.5. ES-SAC ve ES-TD3 ajanları için eklem bazında ortalama mutlak güç değerlerinin karşılaştırılması

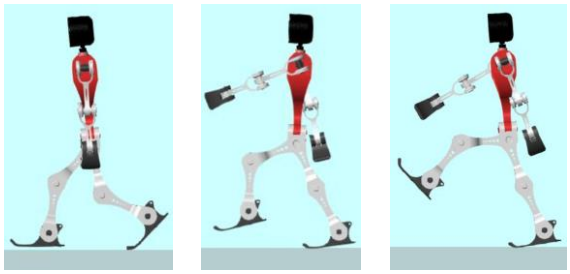
Şekil 4.5 robotun yürüyüş sırasında alt ekstremite eklemlerinde ürettiği tork değerlerini göstermektedir. Daha yüksek tork değerleri ödül fonksiyonunda ceza puanı oluşturduğundan, robot güç verimliliğini korumaya teşvik edilmektedir. Tüm eklemler boyunca ortalama mutlak tork değerleri karşılaştırıldığında, ES-SAC ile denetlenen robot 46.824 Nm kaydederken, ES-TD3 ile denetlenen robot daha yüksek bir ortalama olan 48.936 Nm değerine ulaşmıştır. Bu durum, ES-TD3 ajanının ES-SAC ajanına kıyasla torku daha agresif uyguladığını göstermektedir. Sonuç olarak ES-SAC ajanı, benzer yürüyüş performansını daha düşük tork çıktısıyla sağlayarak daha enerji verimli ve daha yumuşak bir denetim stratejisi sergilemektedir. Sağ ve sol diz eklemi torkları incelendiğinde, ES-SAC ajanının her iki dizde benzer tork değerleri üreterek simetrik denetimi desteklediği görülmüştür. Buna karşılık ES-TD3 ajanı iki diz eklemi arasında belirgin biçimde daha büyük bir tork farkı göstermiştir. Bu kritik eklemlerdeki tork üretim asimetrisi, insansı robotlarda yürüyüş kararlılığını ve simetriyi bozabilmektedir.

Çalışmanın bu kısmına kadar, insansı robot yürüyüş denetimi için önerilen ES-SAC hibrit yöntemi, deterministik ajanlara (ES-TD3 ve ES-DDPG) kıyasla dikkat çekici sonuçlar göstermiştir. ES-SAC ajanı daha yüksek ortalama ödüller elde etmiş ve daha dengeli bir öğrenme süreci sergilemiştir. ES-TD3 ajanı parkuru daha kısa sürede tamamlamış olsa da, bu durum bazı denetim kararsızlıkları ile birlikte ortaya çıkmıştır. Bu nedenle, bu tezin ikinci aşamasında daha insan-benzeri, daha hızlı ve daha yumuşak yürüyüş davranışlarını teşvik edecek biçimde eklemli kol hareketinin modele dahil edilmesi ve dayanıklılığı değerlendirmek amacıyla düz olmayan zeminlerde (örneğin düzensiz karo yüzeyler ve sığ basamaklar) testlerin genişletilmesi sağlanmıştır. Birçok çalışmanın ağırlıklı olarak yüzeysel ödül değeri karşılaştırmalarına dayanmasının aksine (örneğin Fujimoto vd. (2018), Khadka ve Tumer (2018) ve Pourchot ve Sigaud (2019)),

bu çalışmada tork verimliliği, yatay ve düşey sapma ile Q_0 değerleri gibi fiziksel ve davranışsal ölçütler de değerlendirmeye katılarak çok yönlü bir analiz benimsenmiştir. Bu kapsamlı yaklaşım, öğrenme süreçlerinin kararlılığı, enerji tüketimi ve güvenilirliği hakkında daha derin içgörüler sağlamıştır. Ayrıca sonuçlar, ES bileşeninin entegrasyonunun, çözüm uzayında daha geniş keşif yapılmasına olanak tanıyarak ve yerel minimumlardan kaçınmaya yardımcı olarak genel olarak tüm ajanların performansını iyileştirdiğini, böylece daha dayanıklı politikalar ürettiğini göstermiştir. Genel olarak çalışma, öğrenme algoritmalarının yalnızca sayısal performanslarıyla değil, ortaya çıkan davranışların niteliği ve denetim güvenilirliği üzerinden de değerlendirilmesinin önemini vurgulamaktadır.

4.2. Kol Salınım Stratejilerinin Düz ve Engebeli Zeminde İncelenmesi

İlk aşama sonuçlarında ES-SAC yaklaşımının genel performans ve kararlılık açısından daha başarılı sonuçlar vermesi üzerine, ikinci aşamada ayrıntılı inceleme bu yöntem üzerinden sürdürülmüştür. Bu kez analiz kapsamı genişletilerek, yalnızca yürüme başarımı değil, kol salınımının yürüyüş dinamiğine etkisi de değerlendirilmiştir. İkinci aşamadaki deneylerde, görsellerde gösterilen farklı kol kullanım konfigürasyonları altında (örneğin kolların sabit tutulduğu durum ile omuz eklemi üzerinden salınıma izin verilen durumlar) robotun yürüme davranışı karşılaştırılmıştır. Böylece, seçilen en başarılı hibrit denetleyici olan ES-SAC altında kol salınımının denge, koordinasyon ve hareket sürekliliği üzerindeki katkısı sistematik olarak incelenmiştir. Şekil 4.6'da gösterildiği robotun bacakları ve kolları, adım alma sırasında insanın bu iki uzvunda kas kullanımını taklit edecek biçimde tasarlanmış, tork denetimli dönel eklemler (omuz, kalça, diz ve ayak bileği) ile sürülmekte ve kararlı hâlde tutulmaktadır. Omuz eklemi, yalnızca sagittal düzlemde harekete izin veren mekanik bir eksene sahiptir; yani omuz pitch (yanal) ekseni etrafında dönme (ileri-geri salınım) hareketi yapmaktadır (Wang vd. 2017).



(a) (b) (c)

Şekil 4.6. a) Sabit kol konfigürasyonu, her iki kolun tamamen kilitli olduğu ve kol salınımına izin verilmediği durum; **b)** anti-normal kol konfigürasyonu, kolların ipsilateral, yani aynı taraftaki bacak ile aynı fazda salındığı durum; **c)** normal kol konfigürasyonu, insan benzeri karşı fazlı kol-bacak koordinasyonunun sağlandığı durum

Şekil 4.6'da gösterildiği üzere, robotun yürüyüşü üç boyutlu (3B) bir yapı içinde modellenmiş olmakla birlikte, yürüyüş karşılaştırmalarının tutarlı biçimde yapılabilmesi amacıyla bacak ve kol hareketleri sagittal düzlem kinematiği ile sınırlandırılmıştır. Bu kapsamda üç farklı kol konfigürasyonu değerlendirilmiştir. Sabit kol konfigürasyonunda (Şekil 4.6a), kol eklem eksenlerinin tamamı kilitlenerek kolun tüm kinematik zinciri hareketsizleştirilmiş ve böylece kol salınımı tamamen ortadan kaldırılmıştır. Anti-normal (Şekil 4.6b) ve normal (Şekil 4.6c) konfigürasyonlarında ise kol kinematik zinciri sadeleştirilmiş, yalnızca omuz eklemi aktif bırakılmış, omuz dışındaki distal kol eklemleri sabitlenmiştir. Bu iki durumda kol hareketi ayrıca yalnızca omuz ekleminden ileri geri dönme hareketi ile sınırlandırılmıştır. Şekil 4.6b ve Şekil 4.6c arasındaki fark yalnızca kol salınımı ile bacak hareketi arasındaki faz ilişkisine dayanmaktadır. Buna göre anti-normal konfigürasyonda kol salınımı aynı taraftaki bacak hareketi ile aynı fazlı (ipsilateral in-phase) olacak biçimde tanımlanırken, normal konfigürasyonda karşı taraftaki bacak hareketine göre ters fazlı (kontralateral anti-phase/counter-phase) olacak biçimde tanımlanmıştır (Herr ve Popovic 2008).

Çizelge 4.2. Üç yürüyüş konfigürasyonunun, normal, anti-normal ve sabit kol, aktif omuz serbestlik dereceleri, omuz tork aktüasyonu, distal kol eklemlerinin kilitlenme durumu, kol hareket düzlemi kısıtı, uygulanan kol-bacak faz kuralı ve buna karşılık gelen eylem uzayı boyutu açısından tanımları

Koşul	Omuz DOF sayısı (kol başına)	Omuz eklemine uygulanan tork	Distal kol eklemleri	Kol hareket düzlemi	Faz kuralı	Eylem uzayı
Normal	1 (yalnızca omuzun sagittal düzlemde dönmesi)	Evet	Kilitli	Yalnızca sagittal	Kontralateral karşı faz	8 (6 bacak + 2 omuz)
Anti-normal	1 (yalnızca omuzun sagittal düzlemde dönmesi)	Evet	Kilitli	Yalnızca sagittal	İpsilateral aynı faz	8 (6 bacak + 2 omuz)
Sabit	0 (kola ait tüm eklemler kilitli)	Hayır	Kilitli	Yok	Uygulanamaz	6 (yalnızca bacak)

Eğitim sürecinde modelin, çevre ile etkileşimlerini yorumlayabilmesi için temas kuvvetleri, gövdenin uzaydaki konumu ve yönelimi (kinematik parametreler), eklem

açıları, ileri yönlü hareket bileşenleri ile kol, bacak ve gövde arasındaki etkileşimler sürekli olarak izlenmiştir. Eğitim düz bir yüzey üzerinde gerçekleştirilmiştir. Öncelikle aynı düz zemin üzerinde benzetim verileri elde edilmiş, ardından eğitilmiş bu ajanların performansları çevresel değişimlere verdikleri tepkileri ölçmek amacıyla engebeli bir yüzeyde yeniden değerlendirilmiştir. Eğitim ve benzetim süresince ileri yön (y-ekseni), yanal yön (x-ekseni) ve düşey yön (z-ekseni), robotun gövde koordinat takımına göre tanımlanmıştır.

Çizelge 4.2'de Sabit, anti-normal ve normal konfigürasyonlarında kullanılan kola ilişkin modelleme ve denetim ayarları özetlenmiştir. Söz konusu özet, aktif omuz DoF'ları, tork ile doğrudan sürülme (aktüasyon), distal eklem kilitleme durumu, hareket düzlemi kısıtları, faz kuralları ve buna karşılık gelen eylem uzayı boyutunu (sabit için 6, anti-normal ve normal için 8) kapsamaktadır.

Çizelge 4.3'teki düz zemin eğitim performansı incelendiğinde, anti-normal durumunun tohum medyanlarına göre normal ile aynı performans bandında kaldığı, ancak Ortalama Ödül açısından yaklaşık %8-9 daha düşük olduğu görülmektedir. Bu nedenle ödül tabanlı büyüklükler yalnızca normal ve anti-normal arasında karşılaştırılmıştır; sabit kol durumu ise kol ile ilişkili şekillendirme terimleri çıkarıldığı için ödül/Q_0 koşullar arası karşılaştırmalarına dahil edilmemiştir.

Çizelge 4.3. Ortalama ödül ve yürüme kalitesi skoru (Q_0) üzerindeki etkiler; değerler medyan [IQR], n = 5 olarak verilmiştir

Koşul	Ortalama ödül (ortalama [IQR] n = 5)	$\boldsymbol{Q_0}$ (ortalama [IQR] n = 5)
Normal	8760 [8049-8960]	92.3 [82.5-100]
Anti-normal	8020 [7207-8220]	82.5 [78.4-86.3]
Sabit	1540 [1400-1680]	54.4 [51.6-61.0]

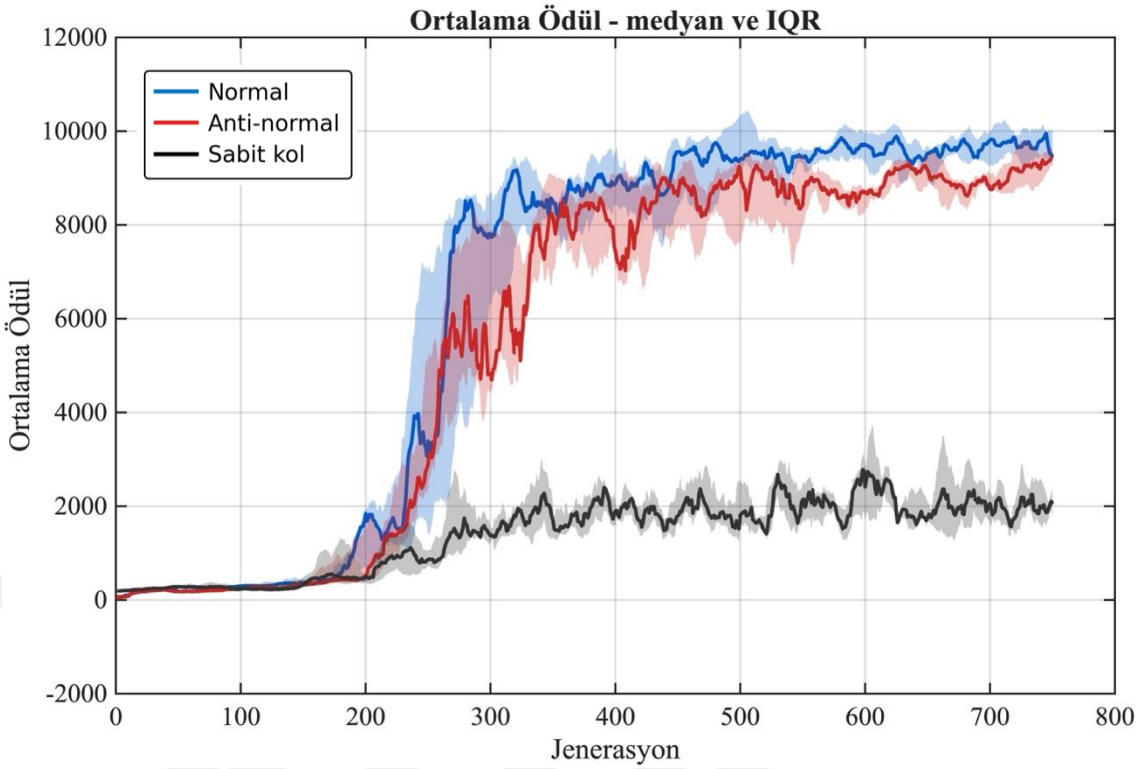
Çizelge 4.4'te verilen yürüyüş istatistiklerinde, anti-normal konfigürasyonu normal konfigürasyonuna kıyasla ileri hızı (V_y) artırmasına rağmen (1.6189'a karşı 1.5423 m/s, yaklaşık %+5.0), yol izleme başarımını hafif ölçüde kötüleştirmektedir; bunun göstergesi olarak $|V_x|$ değeri daha yüksek (0.1747'ye karşı 0.1678 m/s) ve |X| değeri de daha yüksektir (0.2280'a karşı 0.2242 m). Ortalama düşey yer değiştirme dz değeri de anti-normal durumda biraz daha negatiftir (-0.0599'a karşı -0.0545 m). Buna bağlı olarak birim hız başına güç tüketimi ($|P|/V_y$, burada aynı kararlı durum penceresi boyunca ortalama ileri hıza göre normalize edilmiş mutlak toplam eklem gücünü ifade eder ve birimi J/m'dir) normal durumda anti-normal duruma göre daha düşüktür (545.91'e karşı 557.14); bu durum normal konfigürasyonda daha iyi enerji ekonomisine işaret etmektedir. Buna karşılık anti-normal konfigürasyonu birim hız başına daha az tork gerektirmektedir; $|\tau|/V_y$ değeri yaklaşık %9.8 daha düşüktür (189.61'e karşı 210.15). Bununla birlikte, normal konfigürasyonu karşılaştırılabilir hızlarda (1.54'e karşı 1.62 m/s) daha yüksek

getiri ve daha düşük $|P|/V_y$ değeri sağlamaktadır. Sabit kol durumunda ise V_y belirgin biçimde 1.0094 m/s'ye düşmekte, toplam güç ve toplam tork değerleri diğer gruplara göre önemli ölçüde daha düşük kalmaktadır (|P| = 539.07 W; $|\tau|$ = 197.76 N·m). Normalizasyon sonrasında $|P|/V_y$ ve $|\tau|/V_y$ değerleri normal konfigürasyon ile aynı mertebede kalmaktadır (534.03'e karşı 545.91 ve 195.91'e karşı 210.15). Bu nedenle sabit kol durumundaki temel ödünleşim, hızın azalması ve yürüyüş kalitesinin bozulmasıdır; normalize edilmiş maliyetlerdeki küçük farklar bir verimlilik kazanımı olarak yorumlanmamalıdır.

Çizelge 4.4. Düz zemin üzerinde kararlı durum penceresinde hesaplanan yürüyüş metrikleri. Koordinatlar: ileri yön y, yanal yön x, düşey yön z. V_y: ileri hız (m/s); $|V_x|$: ortalama mutlak yanal hız (m/s); dz: ortalama düşey yer değiştirme (m); |X|: ortalama mutlak yanal konum (m); |P|: ortalama mutlak toplam eklem gücü (W); $|P|/V_y$: birim hız başına güç (J/m); $|\tau|$: ortalama mutlak toplam eklem torku (N·m); $|\tau|/V_y$: birim hız başına tork (N·m·s/m)

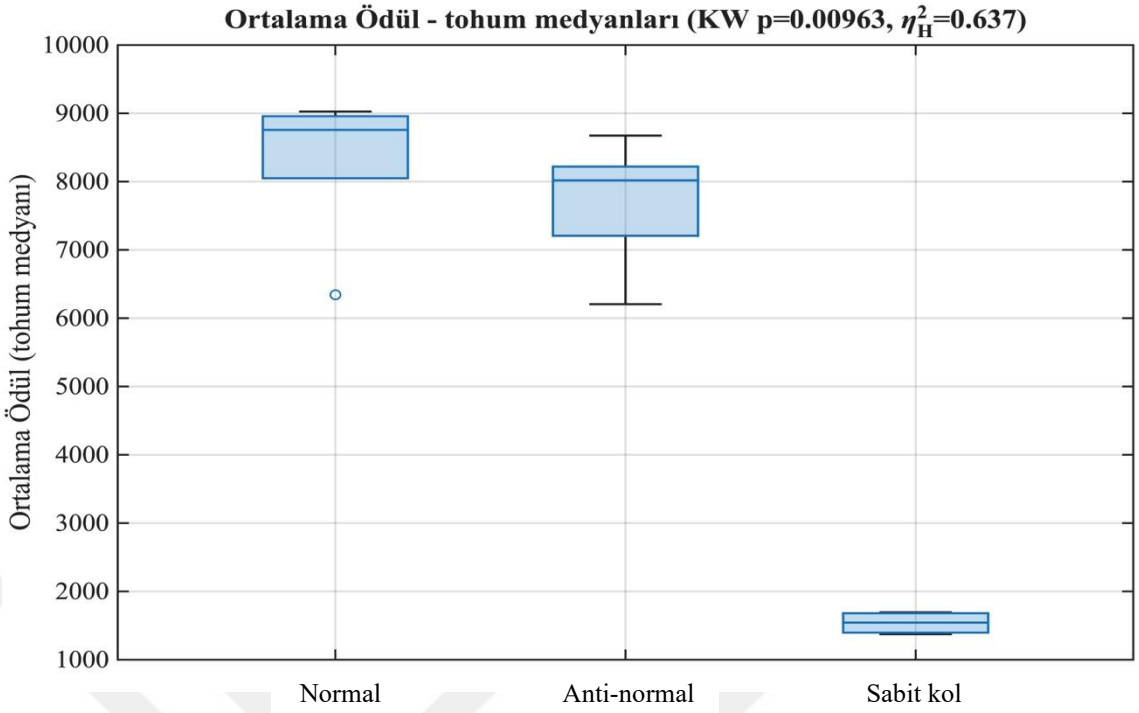
Koşul	$\mathbf{V_y}$	$\mathbf{\lvert V_x \rvert}$	dz	|X|	|P|	$\mathbf{\lvert P \rvert / V_y}$	$\lvert\tau\rvert$	$\mathbf{\lvert\tau\rvert / V_y}$
Normal	1.5423	0.1678	-0.0545	0.2242	841.93	545.91	324.11	210.15
Anti-normal	1.6189	0.1747	-0.0599	0.2280	901.97	557.14	306.97	189.61
Sabit	1.0094	0.1410	-0.0586	0.1810	539.07	534.03	197.76	195.91

Şekil 4.7-4.19 aralığında kullanılan renk şeması şu şekildedir: normal (mavi), anti-normal (kırmızı), sabit kol (siyah). Zaman serisi grafiklerinde yer alan gölgeli bantlar (bulunduğu panellerde) ilgili büyüklüğün tohumlar arası dağılımını temsil etmekte olup medyan etrafındaki IQR değerini göstermektedir. Kutu grafikleri ise her bir tohum için elde edilen özet performans değerlerinin dağılımını sunmaktadır (örneğin tohum medyanları). Raporlanan tüm büyüklükler (örneğin ileri hız, yanal hız, düşey sapma, yanal konum, toplam eklem gücü ve toplam eklem torku gibi) ilgili deney koşulları altında birden fazla rastgele tohum üzerinden elde edilen sonuçlara dayalı olarak karşılaştırmalı biçimde verilmiştir.



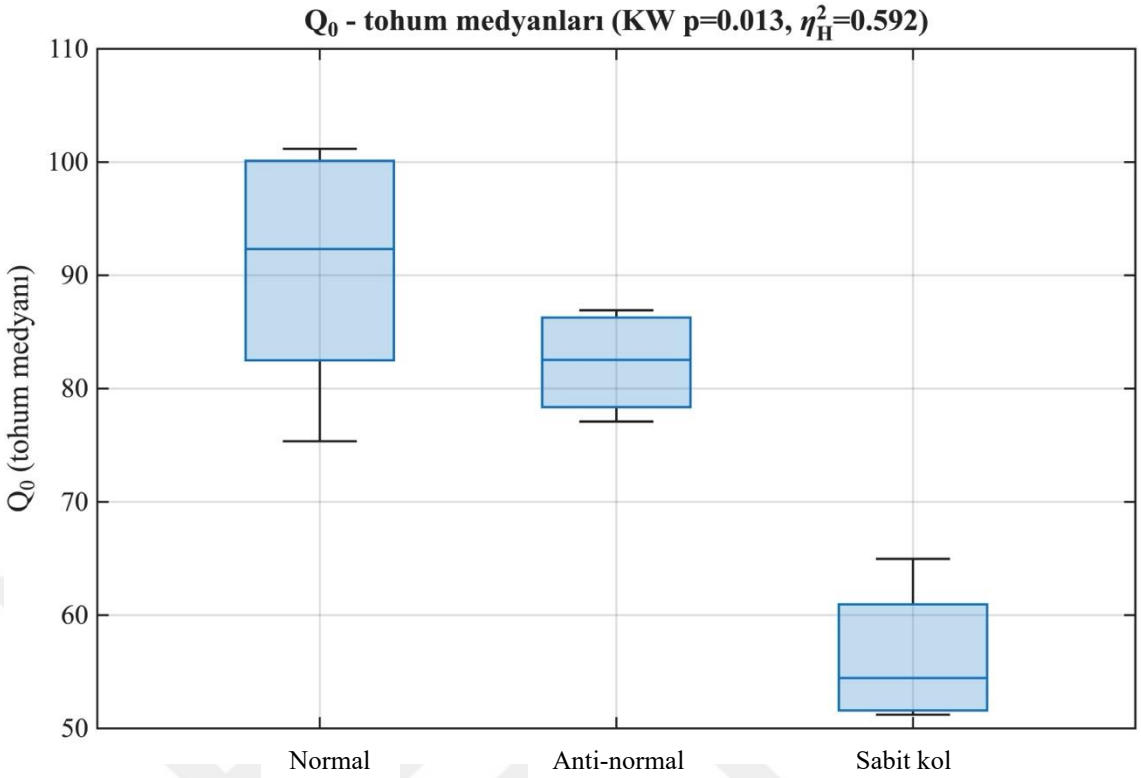
Şekil 4.7. Tüm jenerasyon boyunca ortalama ödülün medyan ve IQR ile değişimi

Şekil 4.7, eğitim sürecinde açık bir yakınsama eğilimine işaret etmektedir. Başlangıçtaki büyüme evresinin ardından tohum-medyan ortalama ödül değeri, ilave iyileşmelerin sınırlı kaldığı bir plato bölgesine girmektedir. Bu kararlı bölgede normal konfigürasyonu daha erken doygunluğa ulaşmakta ve üst ödül bandına yakın seviyede tutarlı biçimde kalmaktadır. Anti-normal konfigürasyonu ise karşılaştırılabilir bir düzeye erişmekle birlikte, normal konfigürasyona göre biraz daha düşük bir tepe değerine ulaşmakta ve plato bölgesinde daha fazla dalgalanma sergilemektedir. Sabit kol konfigürasyonu ise her iki salınımlı stratejiden belirgin biçimde düşük bir ödül bandında seyretmekte ve eğitim boyunca bu farkı kapatamamaktadır; bu durum, kol salınımının yalnızca biyomekanik bir bileşen olmadığını, aynı zamanda öğrenme sinyalinin kalitesini ve yakınsama hızını doğrudan etkilediğini göstermektedir. Ortalama ödül grafiğinde gözlenen bu istikrarlı seyir, verilen eğitim bütçesi ve hiperparametreler altında politikanın büyük ölçüde yakınsadığının bir göstergesi olarak kabul edilmiştir. Ayrıca plato bölgesinde tohum medyanı etrafındaki IQR bandının görece sabit kalması, tohumlar arası değişkenliğin azaldığını göstermekte ve yakınsama yorumunu desteklemektedir.



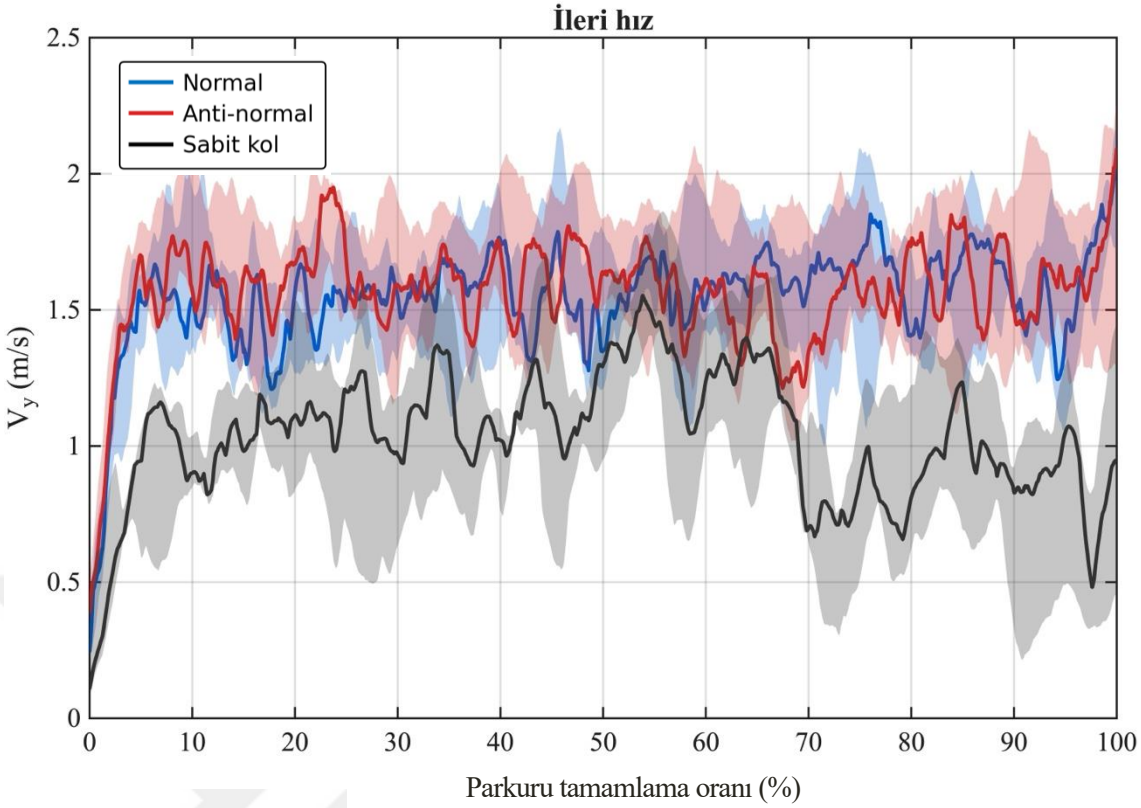
Şekil 4.8. Kol salınım stratejilerine göre ortalama ödül tohum medyanlarının kutu grafiği

Şekil 4.8'de, tohum medyanları normal için 8760 (IQR: 8049-8960) ve anti-normal için 8020 (IQR: 7207-8220) olarak raporlanmıştır. Kruskal-Wallis (KW) testi p = 0.0096 ve η_H^2 = 0.637 sonucunu vermiştir; bu da farkın anlamlı olduğunu göstermektedir. Buna göre eğitim ödülü açısından normal konfigürasyonu anti-normal konfigürasyonundan üstündür. Buna ek olarak, normal konfigürasyonuna ait IQR aralığının (8049-8960) büyük ölçüde anti-normal konfigürasyonunun üst performans bandına karşılık geldiği ve anti-normal medyanının (8020) normal IQR aralığının alt sınırına yakın kaldığı görülmektedir. Bu görünüm, farkın yalnızca tekil tohum sonuçlarından kaynaklanmadığını, tohumlar arası dağılım yapısında da normal lehine bir eğilim bulunduğunu göstermektedir.



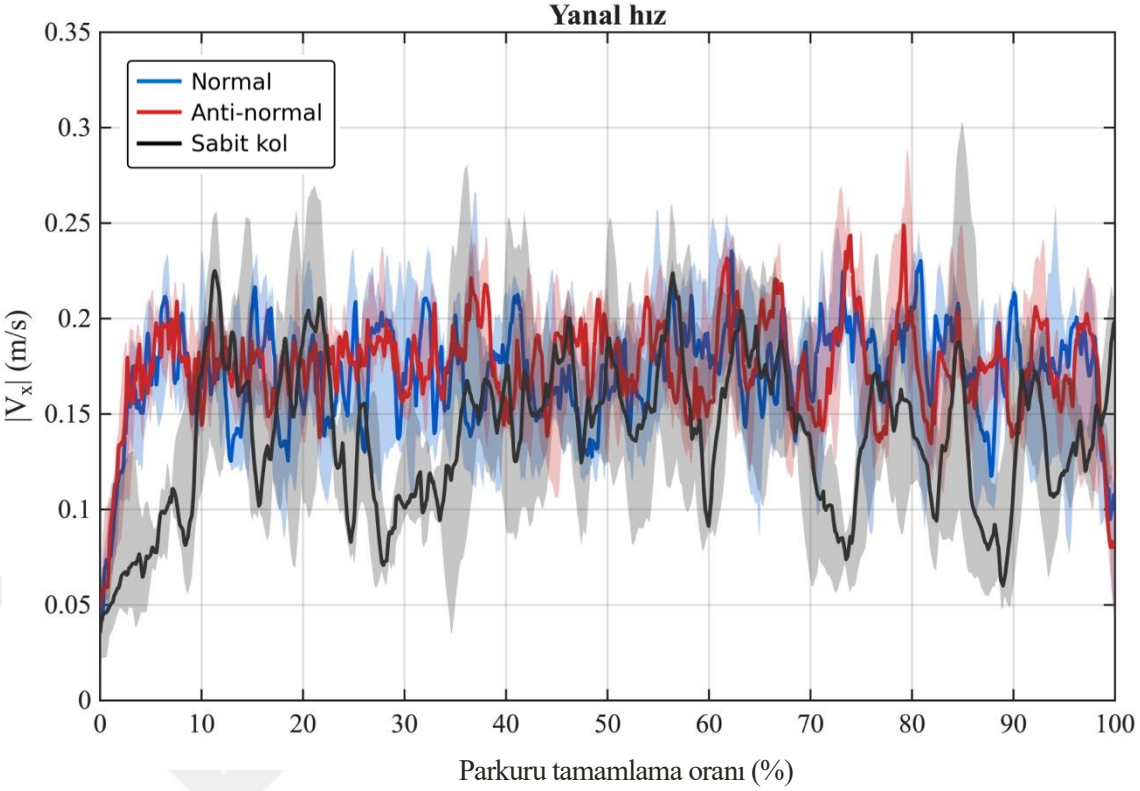
Şekil 4.9. Kol salınım stratejilerine göre Q_0 tohum medyanlarının kutu grafiği

Şekil 4.8'de olduğu gibi Şekil 4.9'da görülen fark yine anlamlıdır (p = 0.013, η_H^2 = 0.592). Q_0 açısından bakıldığında, normal konfigürasyonu başlangıç/erken dönem kalite bakımından da öndedir: normal için 92.3 (82.5-100), anti-normal için 82.5 (78.4-86.3). Ödül tabanlı bu karşılaştırmalar yalnızca normal ve anti-normal arasında yapılmıştır; buna bağlı olarak sabit kol durumu ödül/Q_0 karşılaştırmalarına dahil edilmemiştir. Özellikle normal konfigürasyonunun medyan değerinin anti-normal konfigürasyonunun üst çeyrek düzeyine yakın/üzerinde konumlanması, farkın yalnızca tekil tohum sonuçlarından değil, tohumlar arası merkezi eğilim farkından kaynaklandığına işaret etmektedir. Bu görünüm, normal konfigürasyonunun eğitim sürecinin erken evrelerinde daha yüksek beklenen getiri düzeyiyle ilişkili bir değer fonksiyonu ölçeği ürettiği şeklindeki yorumu desteklemektedir.



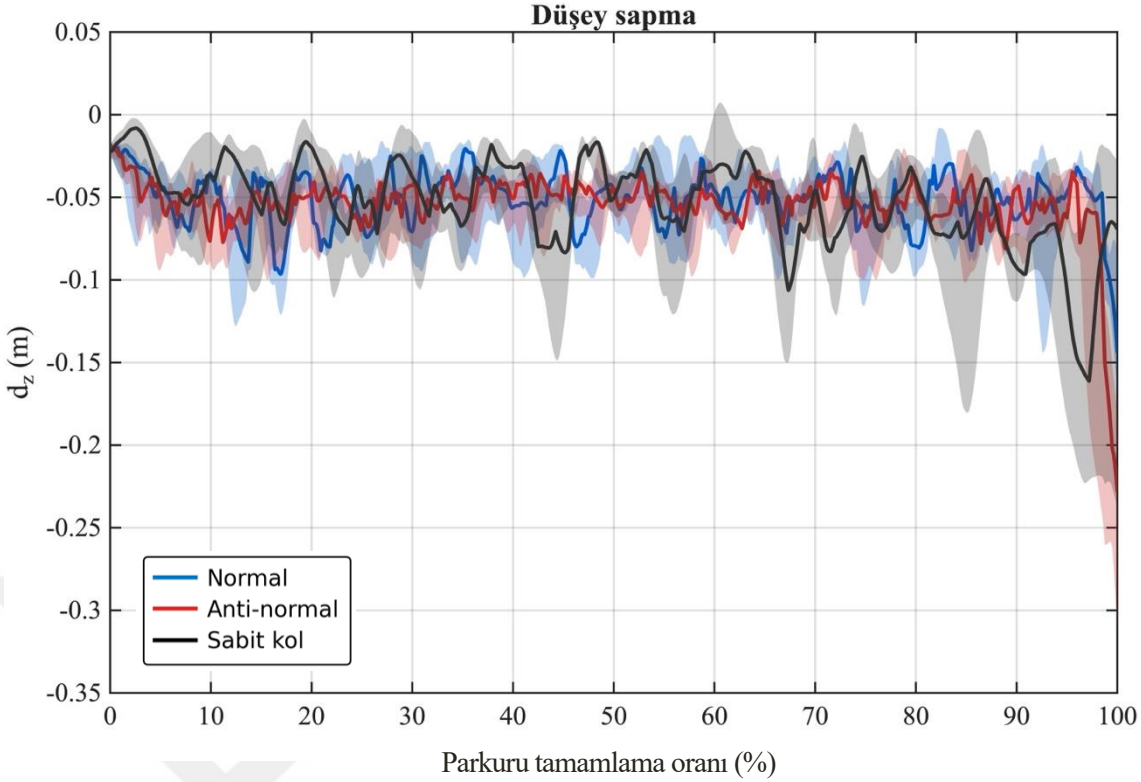
Şekil 4.10. Kol salınım stratejilerine göre ileri hızın normalize zaman boyunca değişimi

Şekil 4.10'de yatay eksen 50 metrelik parkurun ne kadarlık kısmının tamamlandığını yüzde olarak göstermekte (%0 başlangıç, %100 parkur sonu), düşey eksen ise anlık ileri hızı (V_y, m/s) ifade etmektedir. Ortalama ileri hız, anti-normal için 1.6189 m/s ve normal için 1.5423 m/s olup, anti-normal yaklaşık %5 daha hızlıdır. Sabit kol konfigürasyonu ise her iki salınımlı stratejinin belirgin biçimde altında kalmakta ve parkur boyunca daha düşük bir hız bandında seyretmektedir. Grafiğin başlangıç bölgesinde (%0-20 parkuru tamamlama oranı) üç konfigürasyon da benzer bir ivmelenme evresi geçirmekte, ancak yaklaşık %20'den itibaren hız profilleri ayrışmaya başlamaktadır. Anti-normal konfigürasyonunun kararlı durum bölgesinde normal konfigürasyona göre daha yüksek tepe değerlerine ulaştığı, buna karşın normal konfigürasyonunun daha dar bir bant içinde ve daha düzenli bir salınım örüntüsü sergilediği görülmektedir. IQR bantları incelendiğinde, anti-normal konfigürasyonunun tohumlar arası daha geniş bir yayılım gösterdiği ve dolayısıyla hız avantajının tekrarlanabilirlik açısından bir miktar belirsizlik taşıdığı anlaşılmaktadır. Sabit kol konfigürasyonunda ise hem ortalama hızın düşük kalması hem de IQR bandının görece dar olması, kol salınımının olmadığı durumda robotun tutarlı biçimde daha yavaş bir yürüyüş rejiminde kaldığını göstermektedir. Bu bulgular, kol salınım stratejisinin yalnızca ileri hız büyüklüğünü değil, aynı zamanda hız kararlılığını ve tohumlar arası tutarlılığı da doğrudan etkilediğine işaret etmektedir.



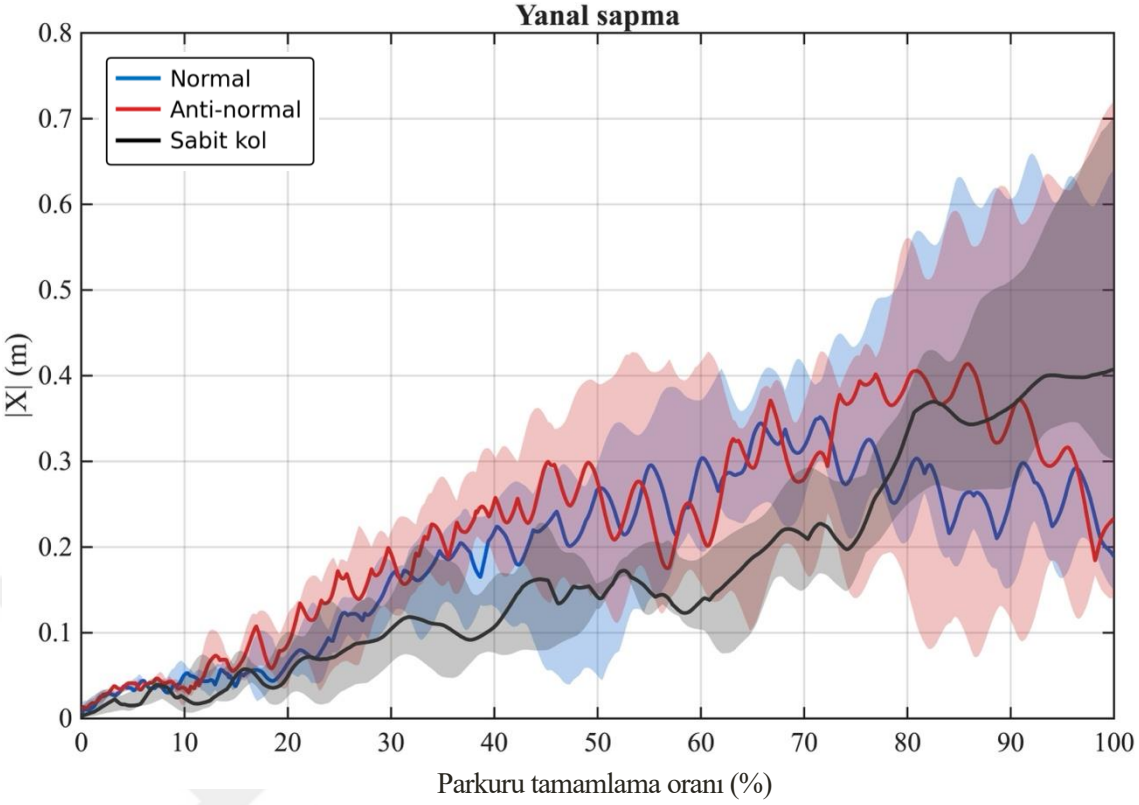
Şekil 4.11. Kol salınım stratejilerine göre yanal hızın normalize zaman boyunca değişimi

Şekil 4.11'de yatay eksen parkuru tamamlama oranını (%), düşey eksen ise ortalama mutlak yanal hızı ($|V_x|$, m/s) göstermektedir. Ortalama değerler birbirine yakındır (normal 0.1678, anti-normal 0.1747 m/s); ancak anti-normal değerinin biraz daha yüksek olması, robotun ileri doğrultuda daha fazla yanal salınım ürettiğine işaret etmekte ve yol izleme doğruluğu açısından normal konfigürasyonunu görece daha tercih edilir kılmaktadır. Sabit kol konfigürasyonu ise parkurun özellikle ilk bölümlerinde (%0-30) diğer iki stratejiye kıyasla daha belirgin yanal hız artışları göstermekte, bu da kol salınımı olmadığında robotun başlangıç dengeleme evresinde daha fazla yanal düzeltme hareketi ürettiğine işaret etmektedir. Kararlı durum bölgesinde üç konfigürasyonun yanal hız profilleri büyük ölçüde örtüşmekle birlikte, IQR bantları incelendiğinde anti-normal konfigürasyonunun tohumlar arası yayılımının normal konfigürasyona göre biraz daha geniş olduğu görülmektedir. Genel olarak yanal hız açısından konfigürasyonlar arasındaki farklar, ileri hıza kıyasla çok daha sınırlı kalmakta; bu durum, yanal dengenin ağırlıklı olarak alt ekstremite denetimi tarafından sağlandığını ve kol salınım stratejisinin bu metrik üzerindeki etkisinin ikincil düzeyde kaldığına işaret etmektedir.



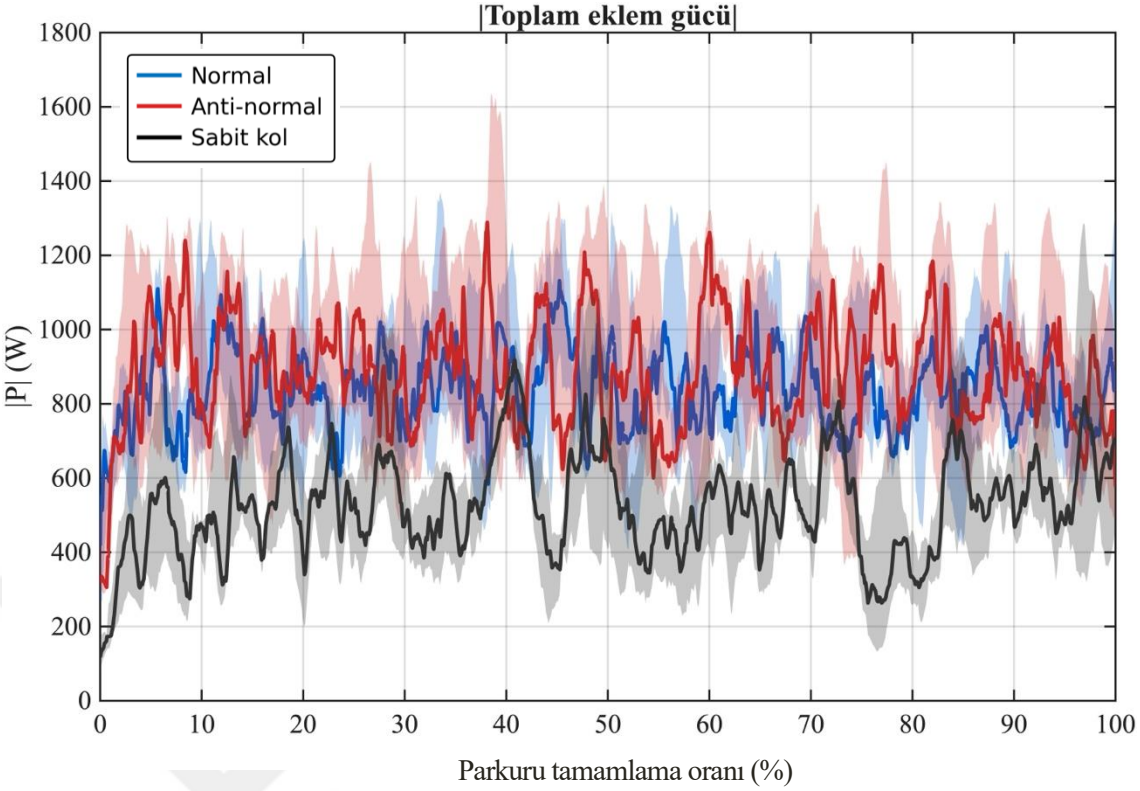
Şekil 4.12. Kol salınım stratejilerine göre düşey sapmanın normalize zaman boyunca değişimi

Şekil 4.12'de yatay eksen parkuru tamamlama oranını (%), düşey eksen ise gövdenin düşey yer değiştirmesini (d_z, m) göstermektedir. Normal kol salınımı için ortalama düşey sapma −0.0545 m, anti-normal için ise −0.0599 m olarak hesaplanmış olup iki salınımlı strateji arasındaki fark oldukça sınırlıdır. Her iki konfigürasyonda da gövde, parkur boyunca sıfır çizgisinin hemen altında dar bir bant içinde seyretmekte ve belirgin bir düşüş eğilimi göstermemektedir. Sabit kol konfigürasyonu ise belirgin biçimde ayrışmaktadır: parkurun ilk yarısında diğer iki stratejiye yakın değerler gözlenmekle birlikte, özellikle %50 sonrasında düşey sapma belirgin biçimde artmakta ve −0.15 ile −0.25 m aralığına kadar inmektedir. Bu durum, kol salınımının olmadığı koşulda robotun gövde yüksekliğini parkurun ilerleyen bölümlerinde giderek daha zor koruduğuna ve düşey denge kapasitesinin azaldığına işaret etmektedir. IQR bantları da bu gözlemi desteklemekte olup sabit kol konfigürasyonunun tohumlar arası yayılımı parkurun ikinci yarısında belirgin biçimde genişlerken, normal ve anti-normal konfigürasyonlarının bantları parkur boyunca dar ve kararlı kalmaktadır. Bu bulgular, kol salınımının düşey eksen stabilitesi üzerinde koruyucu bir rol üstlendiğini ortaya koymaktadır. Grafiğin %100'e yakın son bölümünde bazı konfigürasyonlarda gözlenen ani düşey sapma artışı, robotun 50 metrelik parkurun sonuna ulaşmasının ardından zemin desteğinin sona ermesinden kaynaklanmaktadır. Bu noktada simülasyon veri kaydı devam etmekte, ancak robot serbest düşüşe geçtiğinden bu uç bölgedeki değerler yürüyüş dinamiklerini yansıtmamaktadır. Bu nedenle, raporlanan ortalama düşey sapma değerleri parkurun kararlı durum penceresi içinden hesaplanmış olup son bölümdeki bu geçici etki değerlendirmeye dahil edilmemiştir.



Şekil 4.13. Kol salınım stratejilerine göre yanal sapmanın normalize zaman boyunca değişimi

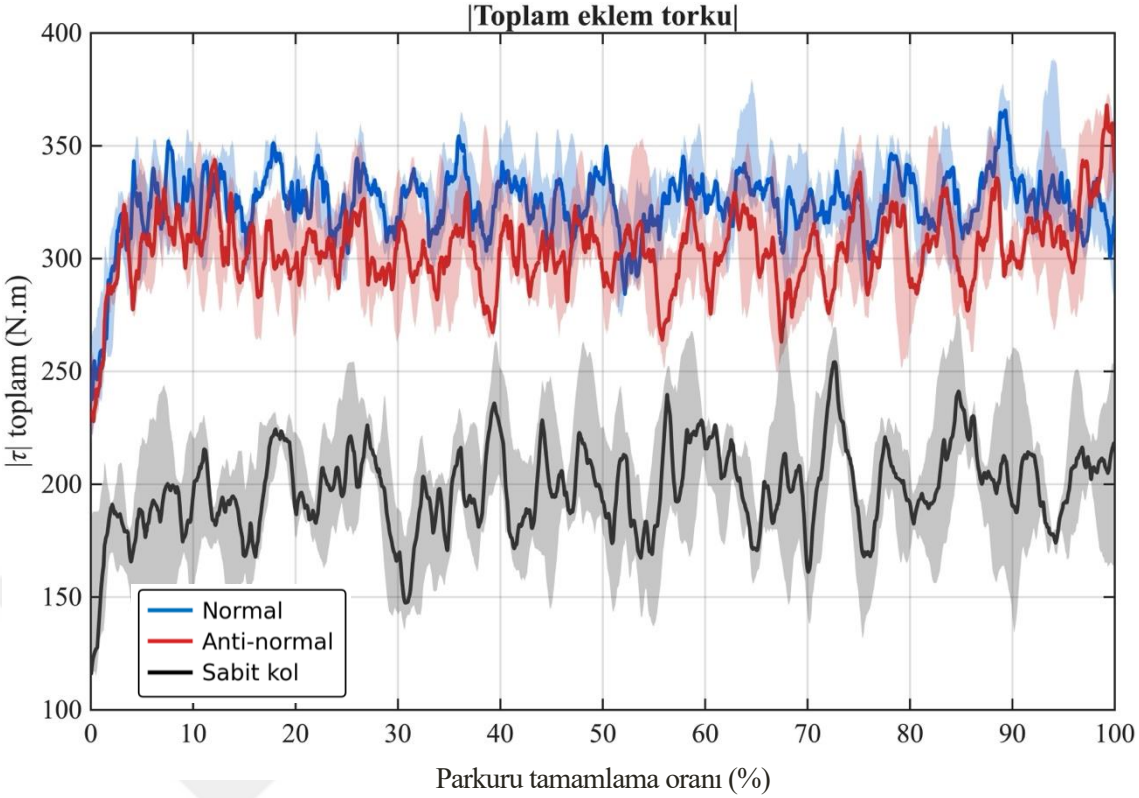
Şekil 4.13'te yatay eksen parkuru tamamlama oranını (%), düşey eksen ise robotun ideal yol doğrultusundan mutlak yanal sapmasını (|X|, m) göstermektedir. Yanal sapma kümülatif bir büyüklük olduğundan, parkur ilerledikçe doğal olarak artmaktadır. Elde edilen ortalama değerler normal konfigürasyonu için 0.2242 m, anti-normal konfigürasyonu için ise 0.2280 m olup, anti-normal konfigürasyonda yanal sapma yaklaşık %1.7 oranında daha yüksektir. Bu bulgu, normal konfigürasyonunun yürüyüş doğrultusunu az da olsa daha iyi koruduğuna işaret etmektedir. Sabit kol konfigürasyonu ise her iki salınımlı stratejiden açık biçimde ayrışmakta ve parkurun sonuna doğru 0.6-0.7 m düzeyine ulaşan belirgin bir yanal kayma sergilemektedir. Bu durum, kol salınımının olmadığı koşulda robotun yanal yöndeki düzeltme kapasitesinin zayıfladığını ve parkur boyunca kümülatif biçimde hedef doğrultudan uzaklaştığını göstermektedir. IQR bantları incelendiğinde, sabit kol konfigürasyonunun tohumlar arası yayılımının parkurun ikinci yarısında belirgin biçimde genişlediği, buna karşılık normal ve anti-normal konfigürasyonlarının daha dar ve tutarlı bir bant içinde kaldığı görülmektedir.



Şekil 4.14. Kol salınım stratejilerine göre toplam eklem gücünün normalize zaman boyunca değişimi

Şekil 4.14'te yatay eksen parkuru tamamlama oranını (%), düşey eksen toplam eklem gücünü (|P|, W); Şekil 4.15'te ise düşey eksen toplam eklem torkunu (|τ| toplam, N·m) göstermektedir. Ortalama toplam eklem gücü anti-normal için 901.97 W, normal için 841.93 W olup anti-normal yaklaşık %7 daha yüksek güç tüketmektedir. Hız ile normalize edildiğinde ($|P|/V_y$) anti-normal için 557.14, normal için 545.91 elde edilmiş olup anti-normal yaklaşık %2 daha maliyetlidir. Dolayısıyla enerji ekonomisi açısından normal konfigürasyonu daha avantajlıdır. Sabit kol konfigürasyonu ise belirgin biçimde düşük bir güç bandında seyretmekle birlikte, bu durum önemli ölçüde düşük ileri hız ve zayıf yol izleme performansıyla birlikte ortaya çıkmaktadır.

Tork açısından ise farklı bir tablo ortaya çıkmaktadır. Anti-normal konfigürasyonunun toplam |τ| değeri 306.97 N·m ile normalin 324.11 N·m değerinin altında kalmaktadır. Hız ile normalize edildiğinde fark daha da belirginleşmekte; $|\tau|/V_y$ anti-normal için 189.61, normal için 210.15 olup anti-normal yaklaşık %9.8 daha düşüktür. Bu sonuç, anti-normal stratejinin daha fazla güç harcamasına rağmen bu gücü daha yüksek hıza daha verimli biçimde dönüştürdüğünü ve tork/hız perspektifinden daha elverişli bir ödünleşim sunduğunu göstermektedir.

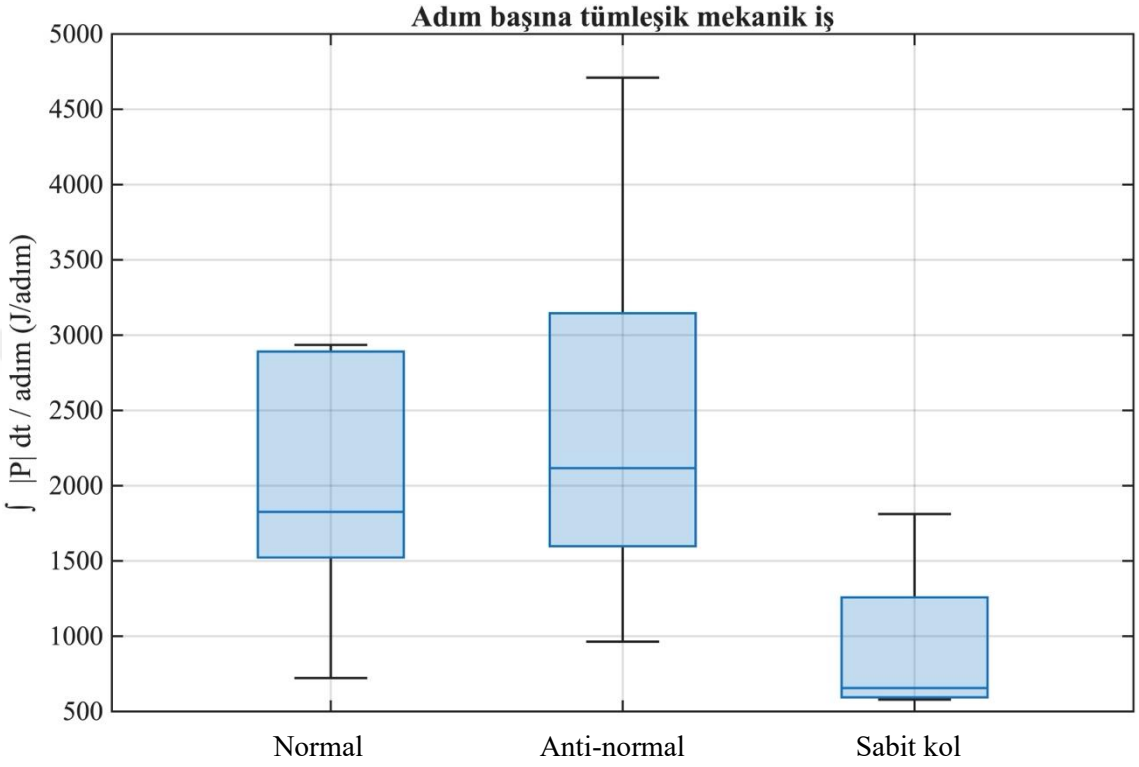


Şekil 4.15. Kol salınım stratejilerine göre toplam eklem torkunun normalize zaman boyunca değişimi

Şekil 4.16-4.19 panellerinde verilen enerji metrikleri, kinematik performans ile enerji ekonomisi arasındaki ödünleşimi nicel olarak ortaya koymaktadır. Kararlı durum penceresi içinde, mutlak eklem gücünün zamana göre integrali kullanılarak adım başına mekanik iş, $W_{step} = \int |P|\, dt$ (J/adım), birim mesafe başına mekanik iş, $W_{dist} = \int |P|\, dt/\Delta y$ (J/m) ve mekanik taşıma maliyeti, $CoT = \int |P|\, dt/(m\, g\, \Delta y)$, hesaplanmıştır. CoT, kararlı durum penceresi üzerinde değerlendirilmekte, vücut ağırlığı ve ileri yöndeki alınan mesafe ile normalize edilmiş mekanik taşıma maliyetini temsil etmekte ve boyutsuz bir büyüklük olarak raporlanmaktadır.

Elde edilen medyan değerler, normal için $W_{\text{step}} \approx 1.83 \times 10^3 J/step$, $W_{\text{dist}} \approx 5.4 \times 10^2 J/m$ ve CoT ≈ 1.1; anti-normal için $W_{\text{step}} \approx 2.12 \times 10^3 J/step$, $W_{\text{dist}} \approx 5.86 \times 10^2 J/m$ ve CoT ≈ 1.2; sabit kol için ise $W_{\text{step}} \approx 6.56 \times 10^2 J/step$, $W_{\text{dist}} \approx 5.15 \times 10^2 J/m$ ve CoT ≈ 1.1 olarak bulunmuştur. Sabit kol konfigürasyonu belirgin biçimde daha düşük V_y ile çalıştığı için, bu gruba ait iş ve CoT benzeri büyüklükler bütünlük amacıyla raporlanmış olup hızdan bağımsız bir enerji ekonomisi üstünlüğü olarak yorumlanmamalıdır. Kutu grafikler, iki kol salınım stratejisi arasındaki enerji temelli ödünleşimi nicel olarak ortaya koymaktadır. Normal ve anti-normal konfigürasyonları karşılaştırılabilir bir lokomosyon rejiminde çalışmakta olup, anti-normal konfigürasyonu daha hızlı olmasına ve birim hız başına daha düşük tork gerektirmesine rağmen, W_{step}, W_{dist} ve CoT açısından daha yüksek mekanik iş göstergeleri sergileme eğilimindedir; bu durum daha yüksek $|P|/V_y$ değeri ile tutarlıdır. Bu nedenle, enerji temelli karşılaştırmalar en anlamlı biçimde normal ve anti-normal arasında yapılmaktadır; çünkü bu iki konfigürasyonda lokomosyon rejimleri birbirine daha yakın olup ödünleşimler daha adil

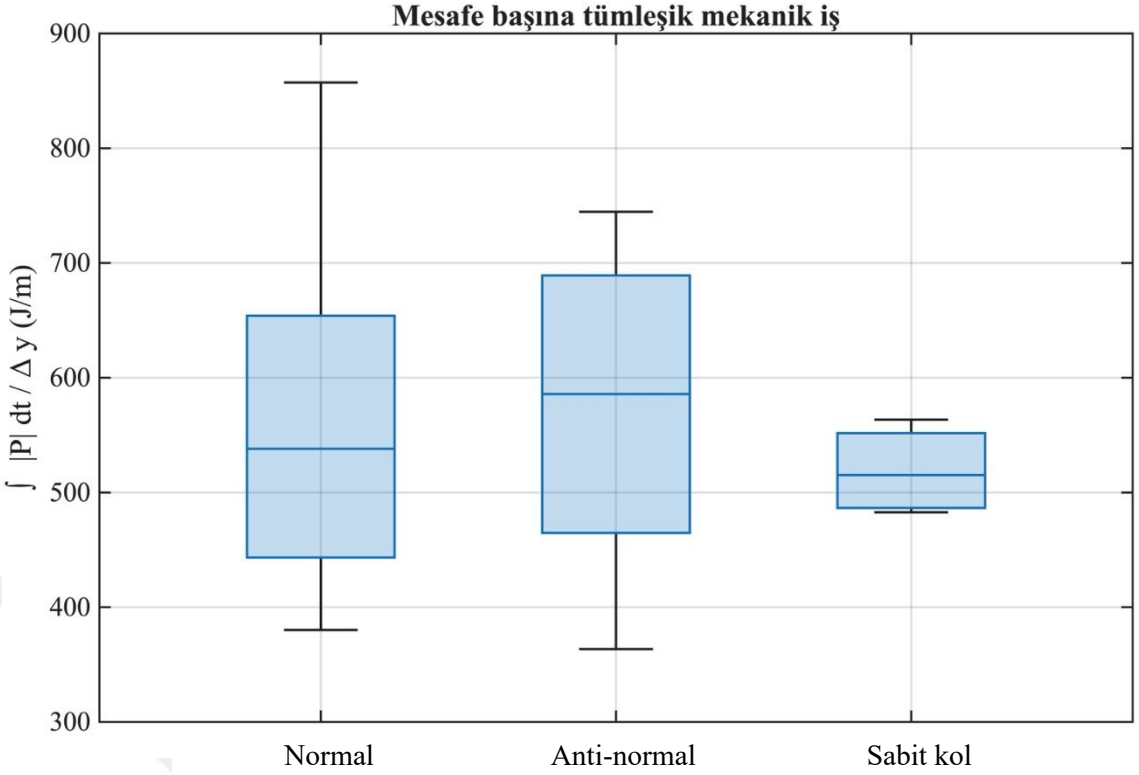
biçimde yorumlanabilmektedir. Sabit kol konfigürasyonu görsel ve analitik bütünlük için sunulmuştur; ancak daha düşük iş/CoT benzeri değerleri, farklı ve daha düşük hızlı bir lokomosyon rejimini temsil ettiği için bire bir bir verimlilik iyileşmesi olarak değerlendirilmemelidir.



Şekil 4.16. Kol salınım stratejilerine göre adım başına tümleşik mekanik iş dağılımı

Şekil 4.16'da verilen adım başına tümleşik mekanik iş ($W_{step} = \int |P|\, dt$, J/adım), robotun tek bir adım atması için tüm eklemlerinin harcadığı toplam mekanik enerjiyi gösterir. Değer ne kadar yüksekse, o adımda eklemler o kadar fazla iş yapmış demektir. Bu metrik, farklı kol salınım stratejilerinin adım düzeyindeki enerji maliyetini doğrudan karşılaştırmaya olanak tanır.

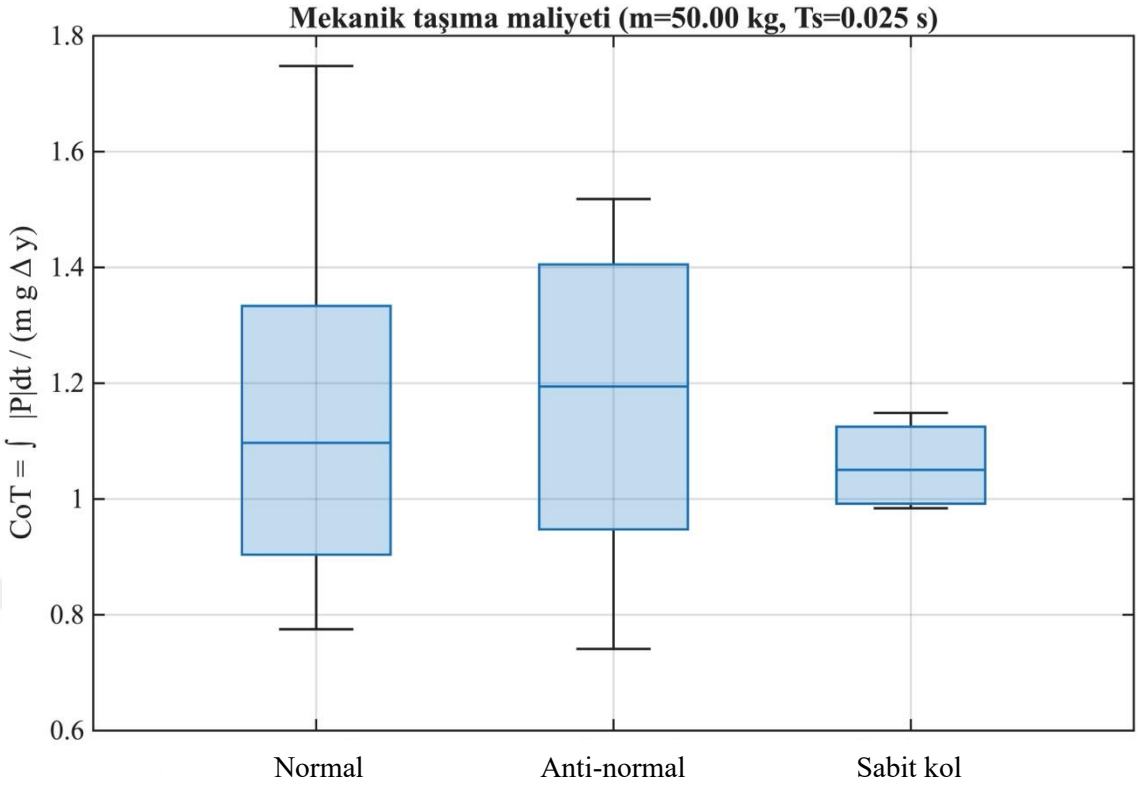
Adım başına tümleşik mekanik iş dağılımı açısından, sabit kol konfigürasyonu diğer iki gruba göre belirgin biçimde daha düşük bir aralıkta yer almaktadır. Normal ve anti-normal grupları ise büyük ölçüde örtüşmektedir; ancak anti-normal grubunda medyan değeri biraz daha yüksektir ve en uç değerler daha yukarıya uzanmaktadır. Bu durum, anti-normal koşulunda adım başına mekanik iş gereksiniminin zaman zaman daha yükseğe çıkabildiğine ve tohumlar arası değişkenliğin daha geniş olduğuna işaret etmektedir. Normal grubu ise benzer bir aralıkta kalmakla birlikte, daha düşük medyan düzeyiyle görece daha tutarlı bir enerji harcama profili sergilemektedir.



Şekil 4.17. Kol salınım stratejilerine göre mesafe başına tümleşik mekanik iş dağılımı

Şekil 4.17'de verilen mesafe başına tümleşik mekanik iş ($W_{dist} = \int |P|\, dt/\Delta y$, J/m) adım başına harcanan toplam enerjinin, o adımda alınan ileri yöndeki mesafeye bölünmesiyle elde edilir. Böylece birim ilerleme mesafesi başına harcanan toplam mekanik enerji elde edilmiş olur. Düşük ileri hıza sahip bir robot, mutlak güç tüketimi sınırlı olsa dahi kat ettiği mesafe kısa kaldığı için normalize edilmiş enerji maliyeti yüksek çıkabilmektedir.

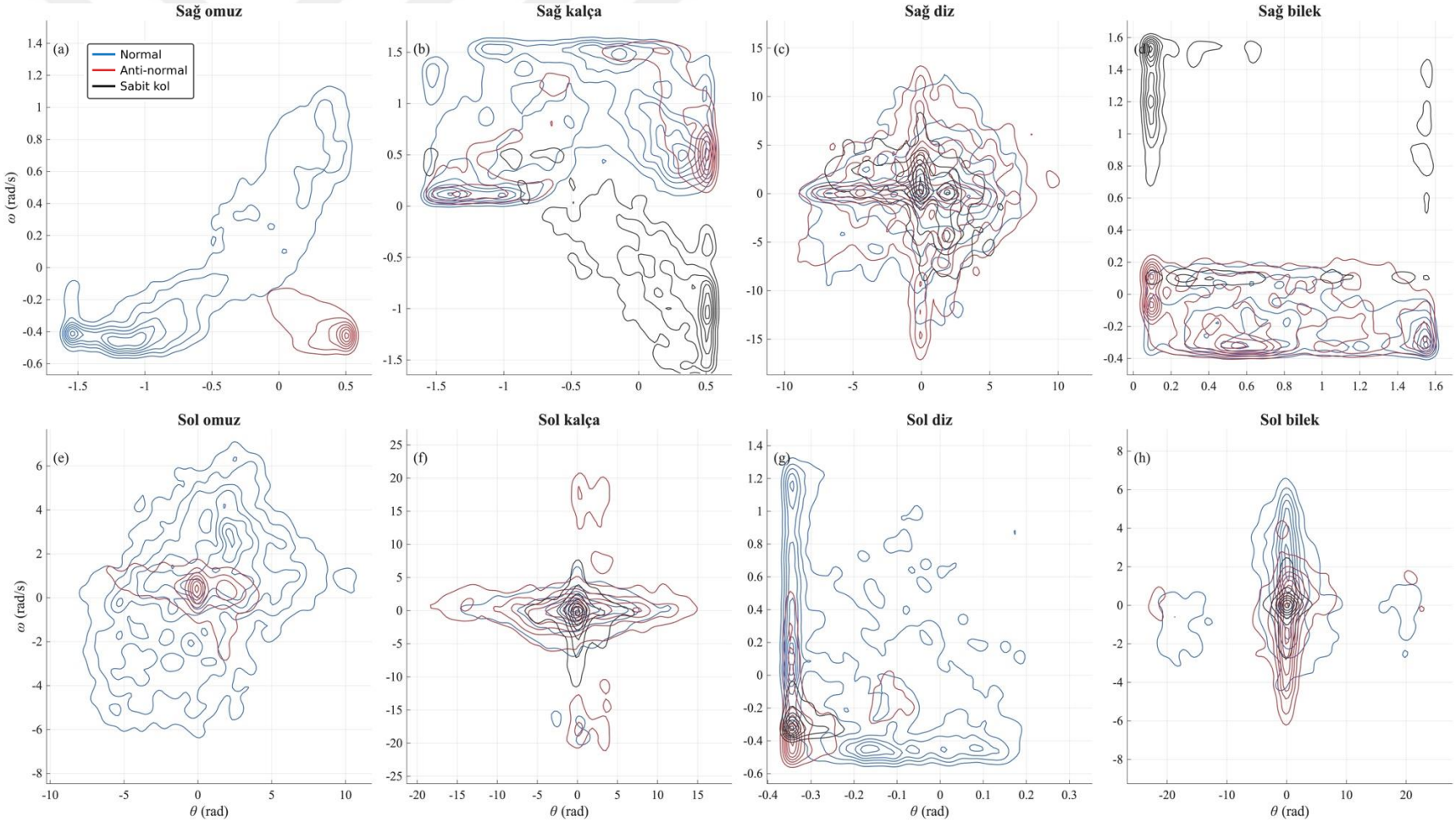
Mesafe başına tümleşik mekanik iş dağılımı metriğinde, normal ve anti-normal gruplarının dağılımlarının yine büyük ölçüde örtüştüğü görülmektedir. Bununla birlikte anti-normal grubunun medyanı normal grubuna göre daha yukarıda konumlanmıştır; bu da birim mesafe başına enerji maliyetinin anti-normal koşulunda daha yüksek olma eğiliminde olduğunu göstermektedir. Sabit kol konfigürasyonu ise dağılım daha dar bir aralıkta toplanmış ve medyan değeri diğer iki grupla aynı mertebede, hatta görsel olarak normal grubuna yakın bir seviyede görünmektedir. Bu durum, mutlak güç/tork düzeyleri düşük olsa bile hızdaki azalma nedeniyle normalize edilmiş enerji maliyetinin belirgin bir üstünlüğe dönüşmediği yönündeki yorumu desteklemektedir.



Şekil 4.18. Kol salınım stratejilerine göre mekanik taşıma maliyeti dağılımı

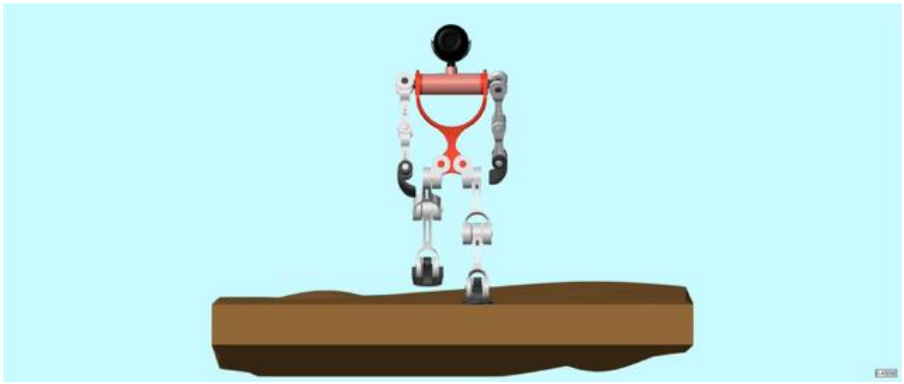
Şekil 4.18'de sunulan mekanik taşıma maliyeti ($CoT = \int |P|\, dt/(m\, g\, \Delta y)$) dağılımı, Birim mesafe başına mekanik işin, robotun ağırlığına ($m \cdot g$) göre normalize edilmiş halidir. Boyutsuz bir büyüklük olduğu için farklı kütlelere sahip robotlar arasında bile karşılaştırma yapmaya uygundur. Düşük CoT değeri, robotun kendi ağırlığını daha verimli taşıdığı anlamına gelir.

Mekanik taşıma maliyeti dağılımı, mesafe başına tümleşik mekanik iş ile tutarlı bir eğilim göstermektedir; çünkü CoT de benzer biçimde enerji tüketiminin ileri yöndeki alınan mesafeye göre normalize edilmiş bir ölçüsüdür. Görsel olarak anti-normal grubunun medyan CoT değeri normal grubundan daha yüksektir ve dağılımı daha geniş bir aralığa yayılmaktadır. Normal grubu daha düşük medyan ve geniş fakat daha aşağı yerleşmiş bir dağılım sergilerken, sabit kol konfigürasyonu daha dar bir aralıkta kümelenmekte ve medyanı diğer gruplarla karşılaştırıldığında aynı mertebede kalmaktadır. Bu görünüm, sabit kol konfigürasyonu normalize edilmiş enerji maliyetinin sınırlı farklar gösterdiğini, buna karşılık normal ve anti-normal arasında enerji ekonomisi açısından normal lehine bir eğilim bulunduğunu göstermektedir.



Şekil 4.19. Üç kol salınım konfigürasyonu için olay hizalamalı limit-çevrim faz portreleri: **a)** sağ omuz; **b)** sağ kalça; **c)** sağ diz; **d)** sağ ayak bileği; **e)** sol omuz; **f)** sol kalça; **g)** sol diz; **h)** sol ayak bileği; omuz eklemlerinde normal ve anti-normal, alt ekstremite eklemlerinde ise sabit kol, normal ve anti-normal konfigürasyonlarına ait yoğunluk konturları gösterilmektedir

Şekil 4.19, eklem açısı ile eklem açısal hızı faz portreleri üzerinden kararlı durum yürüyüş dinamiklerini özetlemektedir. Faz portreleri, tespit edilen ayak temas olayları çevresindeki örneklerin çıkarılması ve bunların adımlar ve tohumlar boyunca olaylara göre hizalanmasıyla oluşturulmuştur. Bu sayede, tekrar tekrar ziyaret edilen kapalı döngüler periyodik bir limit çevrimi temsil ederken, kontur yayılımı adımdan adıma değişkenliği yansıtmaktadır. Omuz eklemleri açısından, normal ve anti-normal konfigürasyonları belirgin biçimde farklı limit-çevrim organizasyonları sergilemektedir. Normal konfigürasyonu daha kompakt ve daha tekrarlanabilir bir omuz çevrimi üretirken, anti-normal konfigürasyonu daha geniş bir doluluk bölgesi ve yer yer artmış bir yayılım göstermekte, bu da adımlar boyunca üst gövde koordinasyonunun daha az tutarlı olduğuna işaret etmektedir. Bu durum, dayatılan koordinasyon desenleriyle uyumludur: normal konfigürasyonu karşı tarafla ters fazlı kol hareketini teşvik ederken, anti-normal konfigürasyonu aynı tarafla aynı fazlı örüntüyü zorlamakta ve bu nedenle gövde ile duruş fazı dinamikleriyle daha güçlü bir bağlaşım oluşturabilmektedir. Alt ekstremite eklemleri açısından ise tüm konfigürasyonlar periyodik lokomosyonu korumaktadır; ancak döngü geometrisi ve yayılım koşullar arasında farklılık göstermektedir. Sabit kol konfigürasyonu genel olarak daha geniş ve daha dağınık kalça, diz ve ayak bileği faz portreleri üretmekte, bu da üst gövdenin kısıtlandığı durumda bacak koordinasyonunda değişkenliğin arttığını düşündürmektedir. Buna karşılık normal konfigürasyonu olasılık kütlesini daha tutarlı bir döngü üzerinde yoğunlaştırma eğilimindedir ve limit çevrimi düzenliliğinin iyileştiğine işaret etmektedir. Anti-normal konfigürasyonu periyodikliğini korumakla birlikte, çevrimin bazı bölümlerinde daha karmaşık bir döngü yapısı ve daha geniş yayılım sunabilmekte, bu da aynı fazlı üst gövde hareketinin getirdiği ek bağlaşım gereksinimleriyle tutarlı görünmektedir.



Şekil 4.20. Engebeli zeminde insansı robot yürüyüşü

Engebeli zemin (Şekil 4.20), düz zeminle aynı boyutlara sahiptir (50 m uzunluk ve 3 m genişlik); ancak düz durumdan farklı olarak tek bir elle tasarlanmış profil yerine, stokastik bir pürüzlü hat olarak üretilmiştir. Daha açık ifadeyle, zemin yüksekliği santimetre ölçeğinde yuvarlatılmış çukur ve tümsekler içeren, rastgele ancak uzaysal olarak korelasyonlu bir yüzeyden oluşturulmuştur. Uzaysal korelasyon uzunluğu yaklaşık 1 m olarak seçilmiştir; böylece eğim değişimleri tekil sivri bozulmalar şeklinde değil, birkaç adım boyunca süren yapılar halinde ortaya çıkmaktadır. Önemli olarak, bu

rastgelelik 50 m'lik hattın tamamına dağıtılmıştır; bozuntular yalnızca yerel bir bölgeyle sınırlı değildir, tüm yürüyüş koridoru boyunca yayılmıştır ve bu nedenle sürekli bir engebeli temas dizisi oluşturmaktadır. Arazi üreteci parametrelendirilebilir yapıdadır (örneğin RMS yükseklik, korelasyon uzunluğu ve çukur/tümsek yoğunluğu) ve rastgele tohum (random seed) ile kontrol edilmektedir; bu sayede aynı istatistiksel özelliklere sahip birden fazla arazi gerçekleşimi, sağlamlık (robust) testleri için üretilebilmektedir. Bu zemin robotu dört açıdan zorlamaktadır: salınım fazında ayak sürtmesi veya zamanlaması hatalı temas riski; duruş fazında kısmi ayak teması ve değişken normal kuvvetler nedeniyle moment dengesinin bozulması; asimetrik eğimler altında yanal sürüklenme ile birlikte sık hız düzeltmeleri ve artan güç/tork gereksinimi; son olarak düz zeminde öğrenilen parametrelerin etkinliğinin değişmesi. Buna bağlı olarak, kol salınım örüntülerinin (normal ve anti-normal) yol takibi ve denge üzerindeki katkısı bu zeminde daha belirgin biçimde gözlemlenebilmektedir.

Çizelge 4.5. Engebeli zemin üzerinde kararlı durum penceresinde hesaplanan yürüyüş metrikleri. Koordinatlar: ileri yön y, yanal yön x, düşey yön z. V_y: ileri hız (m/s); $|V_x|$: ortalama mutlak yanal hız (m/s); dz: ortalama düşey yer değiştirme (m); |X|: ortalama mutlak yanal konum (m); |P|: ortalama mutlak toplam eklem gücü (W); $|P|/V_y$: birim hız başına güç (J/m); $|\tau|$: ortalama mutlak toplam eklem torku (N·m); $|\tau|/V_y$: birim hız başına tork (N·m·s/m). Değerler medyan ∓IQR olarak raporlanmıştır; her politika tohumu beş arazi tohumu üzerinde değerlendirilmiştir

Durum	V_y	$\|V_x\|$	**dz**	**|X|**	**|P|**	$\|P\|/V_y$	$\|\tau\|$	$\|\tau\|/V_y$
Normal	1.5929 ∓0.68	0.1662 ∓0.03	0.0089 ∓0.01	0.4517 ∓0.36	768.33 ∓187	482.35 ∓185	324.28 ∓13	203.58 ∓82
Anti-normal	1.662 ∓0.6	0.1615 ∓0.01	0.0002 ∓0.02	0.2843 ∓0.17	920.95 ∓442	554.13 ∓164	314.39 ∓28	189.17 ∓61
Sabit	1.0644 ∓0.23	0.1165 ∓0.03	-0.047 ∓0.02	0.3294 ∓0.06	522.48 ∓50	490.88 ∓146	198.88 ∓15	186.85 ∓79

Engebeli zeminde değerlendirme protokolünde beş bağımsız arazi gerçeklemesini (terrain seed) kullanılmakta ve her bir arazi tohumu için, her koşulda (normal, anti-normal, sabit) bağımsız olarak eğitilmiş beş politika tohumu test edilmektedir. Adil bir karşılaştırma sağlamak amacıyla, aynı arazi tohumları tüm koşullarda ortak olarak kullanılmıştır. Her bir arazi tohumu için önce beş politika-tohumu denemesi medyan ile birleştirilmiş, ardından beş arazi tohumu boyunca oluşan dağılım medyan ve IQR ile raporlanmıştır. Aynı politikalar düz zeminde (Çizelge 4.4) ve engebeli zeminde (Çizelge 4.5) karşılaştırıldığında iki belirgin eğilim ortaya çıkmaktadır. Birincisi, anti-normal politikasının hız-tork eksenindeki avantajını koruduğu görülmektedir. İkincisi, engebeli zeminde anti-normal politika yol koruma performansını güçlendirirken, normal politika birim hız başına güç maliyeti açısından görece daha güçlü kalmaktadır. Anti-normal her

iki zeminde de daha hızlıdır: düz zeminde normale göre yaklaşık +5.0% daha yüksektir (1.6189'a karşı 1.5423 m/s) ve bu fark engebeli zeminde de yaklaşık +4.3% düzeyinde korunmaktadır (1.662'ye karşı 1.5929 m/s). Yanal birikim (|X|) açısından, düz zeminde normal hafifçe daha iyidir (0.2242 m'ye karşı anti-normal için 0.2280 m; yani anti-normalde |X| yaklaşık +1.7% daha yüksektir); ancak engebeli zeminde anti-normal daha iyi sonuç vermektedir (0.2843 m'ye karşı 0.4517 m; yani normale kıyasla yaklaşık %37 daha düşük |X|). Anlık yanal hız ($|V_x|$) bakımından, anti-normal düz zeminde daha yüksek yanal salınım göstermektedir (0.1747'ye karşı 0.1678 m/s), buna karşılık engebeli zeminde biraz daha düşüktür (0.1615'e karşı 0.1662 m/s). Birim hız başına güç maliyeti ($|P|/V_y$) her iki zeminde de normal için daha düşüktür; düz zeminde fark yaklaşık +2.1% (557.14'e karşı 545.91), engebeli zeminde ise yaklaşık +14.9%'dur (554.13'e karşı 482.35). Birim hız başına tork ($|\tau|/V_y$) avantajı ise anti-normalde kalmaktadır: düz zeminde yaklaşık -9.8% (189.61'e karşı 210.15), engebeli zeminde ise yaklaşık -7.1% (189.17'ye karşı 203.58). Düşey yer değiştirme (dz) büyüklüğü her iki durumda da küçük kalmaktadır: düz zeminde anti-normal biraz daha negatiftir (-0.0599'a karşı -0.0545 m), engebeli zeminde ise her ikisi de sıfıra yakındır ve normal hafifçe daha yüksektir (0.0089'a karşı 0.0002 m).

Collins'in insan yürüyüşü deneyleri (Collins vd. (2009)), doğal (normal) kol salınımının yürüyüşü daha ekonomik hale getirdiğini; buna karşılık kolların kısıtlanmasının (sabit kol, fixed) ve anti-normal (aynı taraflı, aynı fazlı) kol salınımının genel olarak enerji maliyetini artırma eğiliminde olduğunu göstermektedir. Bu çalışmada metabolik enerji doğrudan ölçülmemiş olsa da simülasyon sonuçları mekanik güç/tork tabanlı vekil ölçütler üzerinden değerlendirildiğinde tutarlı bir ayrışma ortaya koymaktadır (normal daha düşük maliyetli bir rejime daha yakın kalmakta, anti-normal daha yüksek maliyet eğilimi göstermekte, sabit kol ise belirgin biçimde farklı bir rejim sergilemektedir). Buna ek olarak, öğrenme eğrileri kol salınım konfigürasyonunun öğrenme verimliliğini doğrudan etkilediğine işaret etmektedir: normal koşulu yüksek ödül bandına daha erken jenerasyonlarda ulaşmakta ve daha hızlı, daha kararlı bir yakınsama sergilemektedir; anti-normal ise karşılaştırılabilir bir düzeye ulaşsa da daha düşük bir medyan/üst sınır ve daha geç bir plato göstermekte, bu da daha zayıf bir öğrenme profiline işaret etmektedir.

Aynı hiperparametreler ve koşullar altında, baskın olarak sagittal düzlem hareketine sahip indirgenmiş DoF uzaysal (3B) bir insansı yürüteç ES-SAC ile eğitilmiş; üç kol stratejisi (normal, anti-normal ve sabit) hem düz hem de engebeli zeminde karşılaştırılmıştır (her koşul için beş bağımsız rastgele tohum). İki salınımlı strateji arasında anti-normal salınım her iki zeminde de daha hızlıdır; buna karşılık birim hız başına torku daha düşüktür (yani eklem çabası verimliliği daha iyidir). Buna karşılık, engebeli zeminde anti-normal salınım yol takibini daha kararlı tutmuştur (daha düşük |X|); buna karşılık normal salınım daha düşük birim hız başına güç maliyetini ($|P|/V_y$) korumuştur. Dolayısıyla zemin düzensiz olduğunda ödünleşim, anti-normal için daha iyi izleme performansı ve normal için daha iyi enerji ekonomisi şeklinde ayrışmaktadır. Collins vd. (2009) çalışmasında kolların sabitlenmesinin metabolik maliyeti yaklaşık %12 artırdığını, aynı fazlı kol salınımının bu maliyeti %26'ya kadar çıkarabildiğini ve kolların sabitlenmesinin düşey reaksiyon momentini yaklaşık %63 artırdığını; buna karşılık normal kol salınımının düşük omuz torku gerektirdiğini göstermiştir. Vekil metriklere dayalı olarak, sabit kol konfigürasyonu esas olarak azalan hız tarafından

belirlenen farklı bir rejim sergilemektedir. Sayısal farkların birebir örtüşmemesi doğaldır; çünkü bu çalışmadaki enerji ölçütü metabolik değil mekaniktir (mutlak eklem gücü toplamı, $|\tau \cdot \omega|$), fakat insan çalışmalarında net metabolik güç raporlanmaktadır. Ayrıca simülatörümüzde bazal metabolizma ile eyleyici/elektronik kayıpları modellenmediğinden, mutlak yüzde değerleri birebir karşılaştırılabilir değildir; yalnızca eğilimler açısından gösterge niteliğindedir.

Özetle, ES-SAC sonuçları kol salınımının yararlı olduğuna ilişkin insan çalışmaları bulgusunu yeniden üretmektedir: hız ve tork verimliliği öncelikli olduğunda anti-normal üstündür; zemin engebeli olduğunda ise anti-normal salınım daha iyi yanal izleme sağlarken, normal salınım daha düşük birim hız başına güç maliyeti ($|P|/V_y$) sunmaktadır.

5. SONUÇLAR

Bu çalışmada, hareketi ağırlıklı olarak sagittal düzlemde gerçekleşen, aktif denetim açısından sekiz serbestlik derecesiyle modellenmiş üç boyutlu bir insansı robot modeli için RL tabanlı yürüme denetimi incelenmiş ve özellikle ES ile desteklenmiş hibrit öğrenme yapılarının yürüyüş kararlılığı, öğrenme başarımı ve enerji ile ilişkili davranışlar üzerindeki etkileri kapsamlı biçimde değerlendirilmiştir.

Çalışmanın ilk aşamasında ES-SAC, ES-TD3 ve ES-DDPG ile bunların saf karşılıkları olan SAC, TD3 ve DDPG ajanları aynı problem ve eğitim koşulları altında ortak değerlendirme ölçütleri kullanılarak karşılaştırılmıştır. Elde edilen bulgular, ES bileşeninin çözüm uzayında daha geniş keşif yapılmasına olanak tanıdığını, yerel minimumlara erken sıkışmayı azaltarak öğrenme sürecinin dayanıklılığını artırdığını ve genel olarak tüm ajanlarda performans iyileşmesine katkı sağladığını göstermiştir. İlk aşama sonuçlarında ES-SAC yaklaşımının genel performans ve kararlılık bakımından daha başarılı bulunması nedeniyle, ikinci aşamadaki ayrıntılı inceleme bu yöntem üzerinden sürdürülmüştür.

Çalışmanın ikinci aşamasında, en başarılı hibrit denetleyici olarak belirlenen ES-SAC yapısı altında kol salınım stratejilerinin insansı yürüyüş dinamiklerine etkisi incelenmiştir. Bu kapsamda normal, anti-normal ve sabit kol stratejileri hem düz zeminde hem de engebeli zeminde, her koşul için beş bağımsız rastgele tohum kullanılarak karşılaştırılmıştır. Böylece yalnızca tek bir eğitim çıktısı değil, tohumlar arası değişkenlik de göz önünde bulundurularak kararlılık, genellenebilirlik ve güvenilirlik hakkında daha güçlü sonuçlara ulaşılmıştır. Bu yaklaşım, öğrenme eğrilerinin yanı sıra yürüyüşe ilişkin fiziksel ve davranışsal ölçütlerin de birlikte değerlendirilmesine olanak tanımış, böylece önerilen denetim yapısının yalnızca sayısal ödül büyüklükleri üzerinden değil, ortaya çıkan hareketin niteliği üzerinden de yorumlanmasını mümkün kılmıştır.

Düz zemin sonuçları, kol salınım stratejisinin yalnızca biyomekanik davranışı değil, aynı zamanda öğrenme verimliliğini de doğrudan etkilediğini göstermiştir. Öğrenme eğrileri açısından normal kol salınımı, yüksek ödül bandına daha erken ulaşan ve daha kararlı yakınsayan yapı olarak öne çıkmıştır. Anti-normal strateji karşılaştırılabilir bir performans düzeyine erişse de daha düşük medyan değerlere ve daha geç plato davranışına sahip olmuş, bu durum daha zayıf bir öğrenme profiline işaret etmiştir. Kararlı durum yürüyüş metriklerinde ise anti-normal stratejinin ileri hız bakımından normal stratejiyi geçtiği, buna karşılık yol izleme başarımı ve bazı denge göstergelerinde görece daha zayıf kaldığı görülmüştür. Sabit kol konfigürasyonu ise hem öğrenme hem de hareket davranışı bakımından diğer iki salınımlı stratejiden ayrışan, daha düşük hız ve farklı dinamik örüntüler sergileyen bir rejim ortaya koymuştur. Bu bulgular, kol hareketinin yalnızca görsel ya da ikincil bir unsur olmadığını, doğrudan yürüyüşün oluşumunu etkileyen işlevsel bir denetim bileşeni olduğunu göstermektedir.

Engebeli zemin sonuçları, düz zeminde öğrenilen politikaların çevresel değişimlere verdiği tepkilerin anlaşılması açısından ayrı bir önem taşımıştır. Bulgular, iki salınımlı strateji arasında anti-normal kol salınımının her iki zeminde de daha yüksek hız ve birim hız başına daha düşük tork ile öne çıktığını, yani eklem çabası verimliliği bakımından avantaj sağladığını göstermiştir. Bununla birlikte zemin düzensizliği arttığında ödünleşim yapısı daha belirgin hâle gelmiştir. Anti-normal salınım engebeli

zeminde daha düşük yanal sapma ile daha iyi yol izleme başarımı sunarken, normal kol salınımı birim hız başına güç maliyeti bakımından daha ekonomik kalmıştır. Bu bulgu, kol salınım stratejileri arasında tüm ölçütler bakımından üstün olan tek bir yapı bulunmadığını; buna karşılık tercih edilmesi gereken stratejinin değerlendirilen performans ölçütüne göre değiştiğini göstermektedir. Başka bir ifadeyle, hız ve tork verimliliği öncelikli olduğunda anti-normal salınım avantaj sağlamakta, enerji ekonomisi öncelikli olduğunda ise normal salınım daha uygun bir seçenek olarak belirmektedir.

Elde edilen sonuçlar, insan yürüyüşünden esinlenen kol kullanımının robotik lokomosyonda anlamlı bir rol oynadığını da göstermektedir. Literatürde normal kol salınımının düşük omuz torku gerektirdiği ve kolların sabitlenmesinin enerji maliyetini artırabildiği belirtilirken, bu tezde metabolik enerji doğrudan ölçülmemiş olsa da mekanik güç ve tork tabanlı vekil ölçütler üzerinden benzer eğilimler gözlenmiştir. Normal kol salınımının daha düşük maliyetli bir rejime yakın kaldığı, anti-normal stratejinin ise daha yüksek hız ve daha iyi tork verimliliği karşılığında daha yüksek enerji eğilimi gösterebildiği görülmüştür. Sabit kol koşulu ise belirgin biçimde farklı bir davranış rejimi oluşturmuştur. Bu yönüyle tez, insan biyomekaniğinden elde edilen bazı temel gözlemlerin, uygun biçimde tanımlanmış öğrenme ve denetim çerçevesi altında insansı robot modellerinde de yeniden üretilebildiğini göstermektedir.

Bu tez çalışmasının bir diğer önemli katkısı, kol salınım stratejilerinin yalnızca nihai yürüyüş çıktıları üzerinden değil, öğrenme süreci üzerindeki etkileriyle birlikte değerlendirilmiş olmasıdır. Bu kapsamda, aynı eğitim düzeni altında normal, anti-normal ve sabit kol stratejilerinin yakınsama eğilimleri, tohumlar arası değişkenlikleri ve değere ilişkin öğrenme sinyalleri birlikte incelenmiştir. Böylece kol koordinasyonunun yürüyüş kararlılığı, denge ve enerji ile ilişkili performans ölçütleri arasındaki ödünleşimleri nasıl şekillendirdiğine ilişkin bütüncül bir çerçeve sunulmuştur. Bu yaklaşım, robotik lokomosyon çalışmalarında yalnızca en yüksek ödül değerine odaklanmanın yeterli olmadığını; ortaya çıkan hareketin kalitesi, güvenilirliği ve çevresel değişimlere karşı dayanıklılığının da en az sayısal öğrenme başarımı kadar önemli olduğunu göstermektedir. Bu nedenle, elde edilen bulgular yalnızca mevcut deney düzeni açısından değil, gelecekte geliştirilecek öğrenme tabanlı insansı yürüme sistemleri için de açıklayıcı bir temel oluşturmaktadır.

Collins vd. (2009) tarafından gerçekleştirilen insan yürüyüşü deneylerinde, doğal kol salınımının yürüyüş ekonomisini desteklediği açık biçimde gösterilmiştir. Aynı çalışmada, kolların sabitlenmesinin metabolik maliyeti yaklaşık %12 artırdığı, aynı fazlı kol salınımının ise bu artışı yaklaşık %26 düzeyine kadar çıkardığı bildirilmiştir. Ayrıca, kolların sabitlenmesi durumunda düşey reaksiyon momentinin de belirgin biçimde yükseldiği rapor edilmiştir. Bu nedenle, Collins vd. (2009) çalışması, kol hareket biçiminin enerji ekonomisi ve yürüyüş dinamiği üzerinde doğrudan etkili olduğunu ortaya koymaktadır.

Bu tez kapsamında elde edilen simülasyon sonuçlarında ise söz konusu yüzdesel değerler birebir yeniden üretilmemiştir. Bunun temel nedeni, burada enerji değerlendirmesinin insan deneylerindeki gibi metabolik ölçümler üzerinden değil, mekanik güç ve tork tabanlı vekil büyüklükler üzerinden yapılmış olmasıdır. Dolayısıyla bu tezde sunulan sonuçlar, insan deneyleriyle doğrudan sayısal eşdeğerlik kurmak amacıyla değil, davranış eğilimlerini karşılaştırmak amacıyla yorumlanmalıdır. Buna

rağmen, simülasyon bulguları genel eğilim bakımından literatürle uyumlu görünmektedir. Vekil metriklere göre normal kol salınımı daha düşük maliyetli bir rejime daha yakın kalırken, anti-normal strateji daha yüksek hız ve daha iyi birim hız başına tork verimliliği sağlamış, sabit kol konfigürasyonu ise belirgin biçimde farklı ve daha sınırlı bir hareket rejimi sergilemiştir. Ayrıca engebeli zeminde anti-normal salınımın daha iyi yanal izleme sağladığı, buna karşılık normal kol salınımının daha düşük birim hız başına güç maliyetini koruduğu görülmüştür. Sonuç olarak bu tezde elde edilen bulgular, insan deneylerinde raporlanan enerji ekonomisi eğilimleriyle genel düzeyde uyum göstermektedir. Ancak burada kullanılan enerji tanımı metabolik değil mekanik olduğu için, karşılaştırmaların yalnızca eğilim düzeyinde değerlendirilmesi gerekmektedir. Gelecek çalışmalar için birkaç doğal genişleme alanı bulunmaktadır. Öncelikle, burada kullanılan hibrit ES-SAC yapısının daha yüksek DoF tam gövde insansı modellere uygulanması ve kol, gövde ile pelvis koordinasyonunun birlikte incelenmesi yararlı olacaktır. İkinci olarak, yalnızca düz ve engebeli zemin değil, eğimli yüzeyler, düzensiz temas profilleri ve dış bozucular içeren daha zorlayıcı çevreler altında genelleme yeteneği araştırılabilir. Üçüncü olarak, mekanik vekil enerji ölçütlerinin yanı sıra metabolik maliyete daha yakın biyolojik ya da aktüatör tabanlı ayrıntılı enerji modelleri entegre edilerek sonuçların fizyolojik yorumu güçlendirilebilir. Son olarak, simülasyondan gerçek sisteme aktarım, ortam parametrelerinin çeşitlendirilmesine dayalı alan uyarlama yaklaşımları ve hibrit model tabanlı denetim yapıları ile birleştirilmiş öğrenme çerçeveleri, bu tezde elde edilen bulguların gerçek robot sistemlerine aktarılması açısından önemli araştırma yönleri olarak değerlendirilmektedir.

Sonuç olarak bu tez, ES ile desteklenmiş DRL yapılarının insansı robot lokomosyonunda etkili ve güçlü bir alternatif sunduğunu göstermiştir. İlk aşamada ES-SAC yaklaşımının karşılaştırılan yöntemler arasında en başarılı ve en kararlı çözüm olarak öne çıktığı, ikinci aşamada ise bu denetleyici altında kol salınım stratejisinin yürüyüş başarımını anlamlı biçimde şekillendirdiği ortaya konmuştur. Normal, anti-normal ve sabit kol stratejileri arasında performans ölçütüne bağlı belirgin ödünleşimler bulunmakla birlikte, genel sonuçlar kol salınımının insansı yürüyüşte pasif bir yan etki değil, denge, kararlılık, enerji ekonomisi ve hareket sürekliliği üzerinde doğrudan etkili yapısal bir unsur olduğunu göstermektedir. Bu yönüyle tez, hem hibrit öğrenme tabanlı robot denetimi literatürüne hem de biyomekanik esinli insansı yürüyüş araştırmalarına özgün ve bütüncül bir katkı sunmaktadır.

6. KAYNAKLAR

Agarwal, R., Schwarzer, M., Castro, P.S., Courville, A. and Bellemare, M.G. 2021. Deep reinforcement learning at the edge of the statistical precipice. Advances in Neural Information Processing Systems, 34: 29304-29320.

Ayyıldız, M. ve Polat, Ö. 2026. Human-like arm swing strategies in ES-SAC humanoid gait: Stability and performance on flat vs rough terrain. Electronic Research Archive, 34(3): 1524-1545. doi: 10.3934/era.2026069.

Bai, H., Cheng, R. and Jin, Y. 2023. Evolutionary reinforcement learning: A survey. Intelligent Computing, 2: 0025. doi: 10.34133/icomputing.0025.

Bakker, B. 2002. Reinforcement learning with long short-term memory. In: Dietterich, T.G., Becker, S. and Ghahramani, Z. (Eds.), Advances in Neural Information Processing Systems 14. MIT Press, Cambridge, pp. 1475-1482.

Bakker, B. 2007. Reinforcement learning by backpropagation through an LSTM model/critic. In: 2007 IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning. IEEE, Honolulu, pp. 127-134.

Bao, L., Humphreys, J., Peng, T. and Zhou, C. 2025. Deep reinforcement learning for bipedal locomotion: A brief survey. Artificial Intelligence Review. doi: 10.1007/s10462-025-11451-z.

Goodfellow, I., Bengio, Y. and Courville, A. 2016. Deep Learning. MIT Press, Cambridge, MA, 800 p.

Bruijn, S.M., Meijer, O.G., Beek, P.J. and Van Dieën, J.H. 2010. The effects of arm swing on human gait stability. Journal of Experimental Biology, 213(23): 3945-3952. doi: 10.1242/jeb.045112.

Callaghan, A., Mason, K. and Mannion, P. 2023. Evolutionary strategy guided reinforcement learning via multibuffer communication. In: Proceedings of the Adaptive and Learning Agents Workshop, AAMAS 2023. Also available as arXiv preprint arXiv:2306.11535.

Chethana, S., Sree Charan, S., Srihitha, V. and Amudha, J. 2023. Humanoid robot gait control using PPO, SAC, and ES algorithms. In: 2023 4th IEEE Global Conference for Advancement in Technology (GCAT). IEEE, Bangalore, pp. 1-7.

Collins, S.H., Adamczyk, P.G. and Kuo, A.D. 2009. Dynamic arm swinging in human walking. Proceedings of the Royal Society B: Biological Sciences, 276(1673): 3679-3688. doi: 10.1098/rspb.2009.0664.

Kaymak, Ç. 2023. İnsansı robotların yürüme becerilerinin derin pekiştirmeli öğrenme algoritmalarıyla geliştirilmesi. Doktora tezi, Fırat Üniversitesi, Elazığ, 233 s.

Denavit, J. and Hartenberg, R.S. 1955. A kinematic notation for lower-pair mechanisms based on matrices. Journal of Applied Mechanics, 22(2): 215-221.

Dong, C. and Li, D. 2024. Adaptive evolutionary reinforcement learning with policy direction. Neural Processing Letters, 56(2): 69. doi: 10.1007/s11063-024-11548-6.

Emmerich, M., Shir, O.M. and Wang, H. 2018. Evolution strategies. In: Martí, R., Pardalos, P.M. and Resende, M.G.C. (Eds.), Handbook of Heuristics. Springer International Publishing, Cham, pp. 89-119.

Fetter, A.L. and Walecka, J.D. 1980. Theoretical Mechanics of Particles and Continua. McGraw-Hill, New York, 608 p.

Fogel, D.B. 2006. Evolutionary Computation: Toward a New Philosophy of Machine Intelligence. 3rd Edition. IEEE Press, Piscataway, NJ, 296 p.

Fujimoto, S., van Hoof, H. and Meger, D. 2018. Addressing function approximation error in actor-critic methods. In: Proceedings of the 35th International Conference on Machine Learning, PMLR, pp. 1587-1596.

Gelbart, M.A., Snoek, J. and Adams, R.P. 2014. Bayesian optimization with unknown constraints. In: Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence, pp. 250-259.

Grondman, I., Busoniu, L., Lopes, G.A.D. and Babuska, R. 2012. A survey of actor-critic reinforcement learning: Standard and natural policy gradients. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42(6): 1291-1307. doi: 10.1109/TSMCC.2012.2218595.

Gu, S., Holly, E., Lillicrap, T. and Levine, S. 2017. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, Singapore, pp. 3389-3396.

Haarnoja, T., Ha, S., Zhou, A., Tucker, G., Tan, J. and Levine, S. 2018a. Learning to walk via deep reinforcement learning. arXiv preprint, arXiv:1812.11103.

Haarnoja, T., Pong, V., Zhou, A., Dalal, M., Abbeel, P. and Levine, S. 2018b. Composable deep reinforcement learning for robotic manipulation. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, Brisbane, pp. 6244-6251.

Haarnoja, T., Zhou, A., Abbeel, P. and Levine, S. 2018c. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proceedings of the 35th International Conference on Machine Learning, PMLR, pp. 1861-1870.

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P. and Levine, S. 2019. Soft actor-critic algorithms and applications. arXiv preprint, arXiv:1812.05905.

Hansen, N. and Ostermeier, A. 2001. Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation, 9(2): 159-195. doi: 10.1162/106365601750190398.

Hashemi, A., Orzechowski, G., Mikkola, A. and McPhee, J. 2023. Multibody dynamics and control using machine learning. Multibody System Dynamics, 58(3): 397-431. doi: 10.1007/s11044-023-09884-x.

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Ali Eslami, S.M., Riedmiller, M. and Silver, D. 2017. Emergence of locomotion behaviours in rich environments. arXiv preprint, arXiv:1707.02286.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D. and Meger, D. 2018. Deep reinforcement learning that matters. In: Proceedings of the AAAI Conference on Artificial Intelligence, 32(1): 3207-3214.

Herr, H. and Popovic, M. 2008. Angular momentum in human walking. Journal of Experimental Biology, 211(4): 467-481. doi: 10.1242/jeb.008573.

Hong, K., Li, Y. and Tewari, A. 2024. A primal-dual-critic algorithm for offline constrained reinforcement learning. arXiv preprint.

Ilboudo, W.E.L., Kobayashi, T. and Sugimoto, K. 2022. Robust stochastic gradient descent with Student-t distribution based first-order momentum. IEEE Transactions on Neural Networks and Learning Systems, 33(3): 1324-1337. doi: 10.1109/TNNLS.2020.3041755.

Islam, R., Henderson, P., Gomrokchi, M. and Precup, D. 2017. Reproducibility of benchmarked deep reinforcement learning tasks for continuous control. arXiv preprint, arXiv:1708.04133.

Issa, A.A. and Aldair, A.A. 2022. Learning the quadruped robot by reinforcement learning (RL). Iraqi Journal for Electrical and Electronic Engineering, 18(2): 117-126. doi: 10.37917/ijeee.18.2.15.

Itahashi, N., Itoh, H., Fukumoto, H. and Wakuya, H. 2024. Reinforcement learning of bipedal walking using a simple reference motion. Applied Sciences, 14(5): 1803. doi: 10.3390/app14051803.

Jung, H., Gu, Z., Zhao, Y., Park, H.-W. and Ha, S. 2025. PreCi: Pretraining and continual improvement of humanoid locomotion via model-assumption-based regularization. arXiv preprint, arXiv:2504.09833.

Kajita, S., Kanehiro, F., Kaneko, K., Fujiwara, K., Harada, K., Yokoi, K. and Hirukawa, H. 2001. The 3D linear inverted pendulum mode: A simple modeling for a biped walking pattern generation. In: Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, Maui, HI, pp. 239-246.

Kasaura, K., Miura, S., Kozuno, T., Yonetani, R., Hoshino, K. and Hosoe, Y. 2023. Benchmarking actor-critic deep reinforcement learning algorithms for robotics control with action constraints. IEEE Robotics and Automation Letters, 8(8): 5055-5062. doi: 10.1109/LRA.2023.3284378.

Kotha, R., Veer, S., Iyer, V. and Lin, H. 2024. Humanoid robot locomotion: A review of control strategies and learning-based approaches. Heliyon, 10(14): e34630. doi: 10.1016/j.heliyon.2024.e34630.

Khadka, S. and Tumer, K. 2018. Evolution-guided policy gradient in reinforcement learning. In: Advances in Neural Information Processing Systems 31. Curran Associates, Red Hook, NY, pp. 1188-1200.

Lee, K., Lee, B.-U., Shin, U. and Kweon, I.S. 2021. An efficient asynchronous method for integrating evolutionary and gradient-based policy search. In: Proceedings of the AAAI Conference on Artificial Intelligence, 35(9): 7816-7824.

Lee, J., Jeon, S., Kim, M., Kim, J. and Hwangbo, J. 2025. Learning humanoid arm motion via centroidal momentum regularized multi-agent reinforcement learning. arXiv preprint, arXiv:2507.04140.

Lehmann, M. 2024. The definitive guide to policy gradients in deep reinforcement learning: Theory, algorithms and implementations. arXiv preprint, arXiv:2401.13662.

Li, L., Xie, Z., Luo, X. and Li, J. 2021. Trajectory planning of flexible walking for biped robots using linear inverted pendulum model and linear pendulum model. Sensors, 21(4): 1082. doi: 10.3390/s21041082.

Li, P., Hao, J., Tang, H., Fu, X., Zheng, Y. and Tang, K. 2024. Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms. arXiv preprint, arXiv:2401.11963.

Li, Z., Peng, X.B., Abbeel, P., Levine, S., Berseth, G. and Sreenath, K. 2024. Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control. arXiv preprint, arXiv:2401.16889.

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D. 2015. Continuous control with deep reinforcement learning. arXiv preprint, arXiv:1509.02971.

Lin, L. and Mitchell, T.M. 1993. Reinforcement learning with hidden states. In: From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior. MIT Press, Cambridge, MA, pp. 271-280.

Lin, Y., Lin, F., Cai, G., Chen, H., Zou, L. and Wu, P. 2024. Evolutionary reinforcement learning: A systematic review and future directions. arXiv preprint, arXiv:2402.13296.

Liu, C.L., Chang, C.C. and Tseng, C.J. 2020. Actor-critic deep reinforcement learning for solving job shop scheduling problems. IEEE Access, 8: 71752-71762. doi: 10.1109/ACCESS.2020.2987820.

Liu, S. 2024. An evaluation of DDPG, TD3, SAC, and PPO: Deep reinforcement learning algorithms for controlling continuous system. In: Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023). Atlantis Press, Paris, pp. 15-24.

Liu, X., Zhang, C., Li, S. and Zhu, J. 2024. Data-driven VVC for unbalanced networks with tuning-friendly deep reinforcement learning. In: 2024 IEEE Power & Energy Society General Meeting (PESGM). IEEE, Seattle, pp. 1-5.

Liu, Z., Li, X., Kang, B. and Darrell, T. 2021. Regularization matters in policy optimization: An empirical study on continuous control. arXiv preprint.

Louette, A., Lambrechts, G., Ernst, D., Pirard, E. and Dislaire, G. 2024. Reinforcement learning to improve delta robot throws for sorting scrap metal. arXiv preprint, arXiv:2406.13453.

Lynch, K.M. and Park, F.C. 2017. Modern Robotics: Mechanics, Planning, and Control. Cambridge University Press, Cambridge, 544 p.

Mann, H.B. and Whitney, D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1): 50-60. doi: 10.1214/aoms/1177730491.

Marchesini, E., Corsi, D. and Farinelli, A. 2021. Genetic soft updates for policy evolution in deep reinforcement learning. In: International Conference on Learning Representations (ICLR).

Masters, D. and Luschi, C. 2018. Revisiting small batch training for deep neural networks. arXiv preprint, arXiv:1804.07612.

McCallum, R.A. 1996. Hidden state and reinforcement learning with instance-based state identification. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 26(3): 464-473. doi: 10.1109/3477.499796.

Millard, M. and Mombaur, K. 2019. A quick turn of foot: Rigid foot-ground contact models for human motion prediction. Frontiers in Neurorobotics, 13: 62. doi: 10.3389/fnbot.2019.00062.

Mitobe, K., Mori, N., Aida, G. and Nasu, Y. 2004. Control of walking robots based on manipulation of angular momentum. Artificial Intelligence Review, 22(3): 189-204.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. 2015. Human-level control through deep reinforcement learning. Nature, 518(7540): 529-533. doi: 10.1038/nature14236.

Müller, N. and Glasmachers, T. 2018. Challenges in high-dimensional reinforcement learning with evolution strategies. In: Parallel Problem Solving from Nature, PPSN XV. Springer, Cham, pp. 640-652.

Nagabandi, A., Clavera, I., Liu, S., Fearing, R.S., Abbeel, P., Levine, S. and Finn, C. 2018. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. arXiv preprint, arXiv:1803.11347.

Obuchowicz, A. 2019. Stable Mutations for Evolutionary Algorithms. Springer, Cham, 228 p.

Olaz, X., Alaez, D., Prieto, M., Villadangos, J. and Astrain, J.J. 2023. Quadcopter neural controller for take-off and landing in windy environments. Expert Systems with Applications, 225: 120146. doi: 10.1016/j.eswa.2023.120146.

Padakandla, S. 2020. A survey of reinforcement learning algorithms for dynamically varying environments. ACM Computing Surveys, 54(6): 1-25. doi: 10.1145/3459991.

Peng, X.B., Berseth, G., Yin, K. and Van De Panne, M. 2017. DeepLoco: Dynamic locomotion skills using hierarchical deep reinforcement learning. ACM Transactions on Graphics, 36(4): Article 41.

Pontzer, H., Holloway IV, J.H., Raichlen, D.A. and Lieberman, D.E. 2009. Control and function of arm swing in human walking and running. Journal of Experimental Biology, 212(4): 523-534.

Pourchot, A. and Sigaud, O. 2019. CEM-RL: Combining evolutionary and gradient-based methods for policy search. In: International Conference on Learning Representations (ICLR).

Roshan, E.S., Kolathaya, S. and Thoppe, G. 2023. Improving sample efficiency in evolutionary RL using off-policy ranking. In: Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), pp. 1188-1196.

Radosavovic, I., Kamat, S., Darrell, T. and Malik, J. 2024. Learning humanoid locomotion over challenging terrain. arXiv preprint.

Radosavovic, I., Xiao, T., Zhang, B., Darrell, T., Malik, J. and Sreenath, K. 2023. Real-world humanoid locomotion with reinforcement learning. Science Robotics, 8(79): eade2256.

Reyad, M., Sarhan, A.M. and Arafa, M. 2023. A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 35(23): 17095-17112. doi: 10.1007/s00521-023-08568-z.

Roaas, A. and Andersson, G.B.J. 1982. Normal range of motion of the hip, knee and ankle joints in male subjects, 30-40 years of age. Acta Orthopaedica Scandinavica, 53(2): 205-208.

Salimans, T., Ho, J., Chen, X., Sidor, S. and Sutskever, I. 2017. Evolution strategies as a scalable alternative to reinforcement learning. arXiv preprint, arXiv:1703.03864.

Schilling, R.J. 1990. Fundamentals of Robotics: Analysis and Control. Prentice Hall, Englewood Cliffs, NJ, 434 p.

Shahriari, B., Swersky, K., Wang, Z., Adams, R.P. and De Freitas, N. 2016. Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1): 148-175. doi: 10.1109/JPROC.2015.2494218.

Shen, X. 2024. Comparison of DDPG and TD3 algorithms in a Walker2D scenario. In: Proceedings of the 2024 International Conference on Computer Engineering and Artificial Intelligence, pp. 148-155.

Singh, R.P., Xie, Z., Gergondet, P. and Kanehiro, F. 2023. Learning bipedal walking for humanoids with current feedback. IEEE Access, 11: 82013-82023. doi: 10.1109/ACCESS.2023.3301175.

Snoek, J., Larochelle, H. and Adams, R.P. 2012. Practical Bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems 25. Curran Associates, Red Hook, NY, pp. 2951-2959.

Soucie, J.M., Wang, C., Forsyth, A., Funk, S., Denny, M., Roach, K.E. and Boone, D. 2011. Range of motion measurements: Reference values and a database for comparison studies. Haemophilia, 17(3): 500-507. doi: 10.1111/j.1365-2516.2010.02399.x.

Spears, W.M., De Jong, K.A., Bäck, T., Fogel, D.B. and De Garis, H. 1993. An overview of evolutionary computation. In: European Conference on Machine Learning. Springer, Vienna, pp. 442-459.

Spong, M.W., Hutchinson, S. and Vidyasagar, M. 2006. Robot Modeling and Control. John Wiley & Sons, New York, 478 p.

Srinivas, A., Laskin, M. and Abbeel, P. 2020. CURL: Contrastive unsupervised representations for reinforcement learning. In: Proceedings of the 37th International Conference on Machine Learning, PMLR, pp. 5639-5650.

Sun, C., Song, M., Cai, D., Zhang, B., Hong, S. and Li, H. 2024. A systematic review of echo state networks from design to application. IEEE Transactions on Artificial Intelligence, 5(1): 23-37. doi: 10.1109/TAI.2022.3225780.

Suri, K. 2022. Off-policy evolutionary reinforcement learning with maximum mutations. In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, pp. 1315-1323.

Suri, K., Shi, X.Q., Plataniotis, K.N. and Lawryshyn, Y.A. 2020. Maximum mutation reinforcement learning for scalable control. arXiv preprint, arXiv:2007.13690.

Sutton, R.S. and Barto, A.G. 2018. Reinforcement Learning: An Introduction. 2nd Edition. MIT Press, Cambridge, MA, 552 p.

Sutton, R.S., McAllester, D., Singh, S. and Mansour, Y. 1999. Policy gradient methods for reinforcement learning with function approximation. In: Advances in Neural Information Processing Systems 12. MIT Press, Cambridge, MA, pp. 1057-1063.

Tang, C., Abbatematteo, B., Hu, J., Chandra, R., Martín-Martín, R. and Stone, P. 2024. Deep reinforcement learning for robotics: A survey of real-world successes. arXiv preprint, arXiv:2408.03539.

Tarasov, D., Surina, A. and Gulcehre, C. 2024. The role of deep learning regularizations on actors in offline RL. arXiv preprint, arXiv:2409.07606.

Toro Icarte, R., Waldie, E., Klassen, T.Q., Valenzano, R., Castro, M.P. and McIlraith, S.A. 2019. Learning reward machines for partially observable reinforcement learning. In: Advances in Neural Information Processing Systems 32. Curran Associates, Red Hook, NY, pp. 1540-1551.

ul Abdeen, Z., Zhang, X., Gill, W. and Jin, M. 2024. Enhancing distribution system resilience: A first-order meta-RL algorithm for critical load restoration. In: 2024 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, pp. 129-134.

Umberger, B.R. 2008. Effects of suppressing arm swing on kinematics, kinetics, and energetics of human walking. Journal of Biomechanics, 41(11): 2575-2580. doi: 10.1016/j.jbiomech.2008.05.024.

Wang, Q., De Baets, L., Timmermans, A., Chen, W., Giacolini, L., Matheve, T. and Markopoulos, P. 2017. Motor control training for the shoulder with smart garments. Sensors, 17(7): 1687. doi: 10.3390/s17071687.

Wang, X., Guo, W., Yin, S., Zhang, S., Zha, F., Li, M., Wang, P., Li, X. and Sun, L. 2025. Walking control of humanoid robots based on improved footstep planner and whole-body coordination controller. Frontiers in Neurorobotics, 19: 1538979. doi: 10.3389/fnbot.2025.1538979.

Wang, Z., Pei, Y. and Li, J. 2023. A survey on search strategy of evolutionary multi-objective optimization algorithms. Applied Sciences, 13(7): 4643. doi: 10.3390/app13074643.

Wierstra, D., Schaul, T., Glasmachers, T., Sun, Y., Peters, J. and Schmidhuber, J. 2014. Natural evolution strategies. Journal of Machine Learning Research, 15: 949-980.

Wu, F., Gu, Z., Wu, H., Wu, A. and Zhao, Y. 2023. Infer and adapt: Bipedal locomotion reward learning from demonstrations via inverse reinforcement learning. arXiv preprint, arXiv:2309.16074.

Wu, Z., Lu, J., Chen, Y., Liu, Y., Zhuang, Y. and Hu, L. 2025. STRIDE: Automating reward design, deep reinforcement learning training and feedback optimization in humanoid robotics locomotion. arXiv preprint, arXiv:2502.04692.

Xi, L., Wu, J., Xu, Y. and Sun, H. 2021. Automatic generation control based on multiple neural networks with actor-critic strategy. IEEE Transactions on Neural Networks and Learning Systems, 32(6): 2483-2493. doi: 10.1109/TNNLS.2020.3006080.

Zhang, K., Yang, Z. and Başar, T. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. In: Vamvoudakis, K.G., Wan, Y., Lewis, F.L. and Cansever, D. (Eds.), Handbook of Reinforcement Learning and Control. Springer International Publishing, Cham, pp. 321-384.

Zhang, X., Wang, X., Zhang, L., Guo, G., Shen, X. and Zhang, W. 2024. Achieving stable high-speed locomotion for humanoid robots with deep reinforcement learning. arXiv preprint, arXiv:2409.16611.

Zhang, Y., Xie, J., Du, X., Sun, H., Wang, S. and Hashimoto, K. 2024. Biped robot terrain adaptability based on improved SAC algorithm. In: Larochelle, P. and McCarthy, J.M. (Eds.), Proceedings of MSR-RoManSy 2024. Springer Nature Switzerland, Cham, pp. 93-104.

Zhang, Y. and Chen, P. 2023. Path planning of a mobile robot for a dynamic indoor environment based on an SAC-LSTM algorithm. Sensors, 23(24): 9802. doi: 10.3390/s23249802.

Zhu, Y., Wang, Z., Zhu, Y., Chen, C. and Zhao, D. 2024. Discretizing continuous action space with unimodal probability distributions for on-policy reinforcement learning. arXiv preprint, arXiv:2408.00309

ÖZGEÇMİŞ

Mustafa AYYILDIZ

ÖĞRENİM BİLGİLERİ

Doktora 2022-2026	Akdeniz Üniversitesi Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği Anabilim Dalı, Antalya
Yüksek Lisans 2019-2021	Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği Anabilim Dalı, Isparta
Lisans 2014-2019	Akdeniz Üniversitesi Mühendislik fakültesi, Elektrik-Elektronik Mühendisliği, Antalya

MESLEKİ VE İDARİ GÖREVLER

Öğretim Görevlisi 2023-Devam Ediyor	Akdeniz Üniversitesi Teknik Bilimler Meslek Yüksekokulu, Antalya
Elektrik-Elektronik Mühendisi 2020-2023	Türkiye Radyo ve Televizyon Kurumu TRT Vericiler Dairesi Başkanlığı, Antalya

ESERLER

Uluslararası hakemli dergilerde yayımlanan makaleler

1- Ayyıldız, M. ve Polat, Ö. 2026. Human-like arm swing strategies in ES-SAC humanoid gait: Stability and performance on flat vs rough terrain. Electronic Research Archive, 34(3): 1524-1545. doi: 10.3934/era.2026069.

2- Ayyıldız, M. ve Tilki, U. 2023. Adaptive sliding mode based fault tolerant control of wheeled mobile robots. Automatika, 64(3): 467-483. doi: 10.1080/00051144.2023.2190866.

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler

1- Ayyıldız, M. ve Tilki, U. 2021. Fault tolerant control of wheeled mobile robots using adaptive sliding mode controller. 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), ss. 1-6, Kocaeli, Turkey. doi: 10.1109/INISTA52262.2021.9548505.

2- Ayyıldız, M. ve Tilki, U. 2021. Adaptive sliding mode control of mobile robot and comparison with PID controller. 2nd International Symposium on Applied Sciences and Engineering, ss. 494-498, 7-9 Nisan, Erzurum, Türkiye.