# Pekiştirmeli Öğrenme Tabanlı Dört Pervaneli Uçuş Kontrolcü Tasarımı
# A Reinforcement Learning Based Quadcopter Flight Controller Design

*Burhan Burak Akman*[1], *Banu Kabakulak*[2], *Şeref Naci Engin*[1,3]

[1]Aviyonik Mühendisliği LÜ Programı
Yıldız Teknik Üniversitesi, İstanbul
burhanburakakman@gmail.com

[2]Endüstri Mühendisliği Bölümü
İstanbul Bilgi Üniversitesi, İstanbul
banu.kabakulak@boun.edu.tr

[3]Kontrol ve Otomasyon Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, İstanbul
nengin@yildiz.edu.tr

## Özetçe

Dört pervaneliler günümüzde askeri alanların gözetlenmesi, tarım arazilerinin ilaçlanması, afet durumlarında telekomünikasyon hizmeti sağlanması gibi birçok farklı alanda yaygın olarak kullanılan İnsansız Hava Araçlarıdır (İHA). Bir dört pervanelinin hareketi birbirinden bağımsız dönebilen dört motorla sağlanır. Bir dört pervanelinin belirlenen yörüngeye azami uyumlu uçabilmesi için motor dönme hızlarını gerçek zamanlı kontrol edebilen dayanıklı uçuş kontrolcülerinin tasarlanması gerekmektedir. Bu doğrultuda, literatürde oransal-integral-türevsel (PID) denetleyici, doğrusal karesel düzenleyici (LQR), model öngörülü kontrol (MPC) gibi birçok denetleyici önerilmiştir. Bu denetleyicilerin başarımı seçilen model parametrelerine göre değişkenlik göstermektedir.

Bu çalışmada, dört pervaneli bir İHA'nın verilen bir yörüngeyi yüksek hassasiyetle takip edebilmesi için pekiştirmeli öğrenme tabanlı bir uçuş kontrolcüsü önerilmiştir. Derin belirgin yöntem eğilimi (DDPG) yöntemi ve MATLAB simülasyon ortamını beraber kullanan kontrolcü, aktör-eleştirmen sinir ağlarının tekrarlı çözümler üretmesiyle parametreleri öğrenmektedir. Geliştirilen kontrolcünün farklı referans yörüngeleriyle yapılan testlerinde literatürde sunulan çalışmalara kıyasla tatmin edici performans sergilediği gözlemlenmiştir.

## Abstract

Quadcopters are Unmanned Aerial Vehicles (UAV) which find wide application areas such as surveillance of military zones, spraying the farmlands in agriculture, and providing telecommunication service in case of a disaster. The four independently operating rotors govern the motion of a quadcopter. High-fidelity flight of a quadcopter to a reference trajectory requires the design of a robust real-time flight controller for the rotor speeds. In this direction, many controllers are proposed in the literature such as Proportional-Integral-Derivative (PID) controller, Linear Quadratic Regulator (LQR), and Model Predictive Control (MPC). The performance of these controllers significantly depends on the selected model parameters.

In this work, we develop a reinforcement learning-based trajectory controller with high tracking precision for a quadcopter. This controller couples the Deep Deterministic Policy Gradient (DDPG) algorithm with the MATLAB simulation environment and learns the model parameters from the iterative solutions of actor-critic neural networks. The proposed controller reveals promising performance in tracking various reference trajectories compared to the ones reported in the literature.

## 1. Giriş

*Dört pervaneli* İnsansız Hava Araçları (İHA'lar) basit bir tasarıma sahiptir ve aerodinamik olarak verimli olduklarından günümüzde keşif, gözetleme, tarım, haritalama, fotoğrafçılık, kargo taşımacılığı, savunma ve güvenlik gibi birçok alanda yaygın bir şekilde kullanılmaktadır [1].

Bir dört pervaneli birbirinden bağımsız dönen dört motor (*rotor*) sayesinde itme kuvveti üretir. Tüm rotorlar aynı hızla saat yönünde dönerse kalkış, tersi yönde dönerse iniş gerçekleşir. Dört pervanelinin havada dengede (*hover*) kalabilmesi için aynı eksen üzerindeki iki rotorun saat yönünde, diğer iki rotorun ters istikamette aynı hızla dönmesi gerekir. Kalkış ve inişte tüm rotorlar aynı yönde itki kuvveti oluştururken, denge durumunda dört pervaneliye etki eden rotor momentleri sıfırlanmıştır. Rotorların farklı hızlarda dönmeleri dört pervanelinin farklı eksenler etrafında dönmesini sağlar. Böylece bir dört pervaneli, rotor hızlarının kontrol edilmesiyle hem ötelemeli (*translational*)

hem de dairesel (*rotational*) manevralar yapabilir [2].

Rotor hızların belirlenmesi için literatürde çeşitli kontrol yaklaşımları benimsenmiştir. Birinci yaklaşım, bir dört pervanelinin uçuş dinamiklerinin diferansiyel denklemlerle matematiksel modellenmesine dayanır. Bu geleneksel yaklaşımda elde edilen model doğrusal değildir, ancak küçük açı varsayımı gibi özel durumlar için doğrusallaştırılabilir [3]. Kontrol teorisi araçlarını kullanan böyle bir kontrolcünün her sistem için tasarlanması ve en iyi parametrelerinin belirlenmesi kolay olmamakla beraber sistemin kararlılığı ve hata başarımı hakkında önemli bilgiler sunması açısından değerlidir [4].

Literatürde dört pervaneli dinamik modelini kullanan en yaygın yöntem oransal-integral-türevsel (*Proportional-Integral-Derivative*, PID) kontrolcüdür. Gerçek zamanlı kolay uygulanabilir bir yöntem olan PID, ölçülen hataları belirli katsayılarla kontrol komutlarına dahil eder. PID'nin hata başarımı bu katsayılara bağlıdır ve çeşitli yöntemlerle seçilebilir [5].

Model öngörülü kontrol (*Model Predictive Control*, MPC) yöntemi sistemin dinamik modelini kullanır ve belli bir aralıkta sistemin nasıl davranacağını gerçek zamanlı bir karesel eniyileme yöntemiyle tahmin eder. Bu tahminler uygulanacak kontrol komutlarının birden fazla kısıtı gürbüz bir şekilde karşılayabilmesine olanak tanır. Dört pervaneli yörünge kontrolü için MPC yöntemini bulanık mantık ile harmanlayan doğrusal olmayan bir otopilot önerilmiştir [6]. Doğrusal olmayan MPC yöntemlerinin başarılı yörünge takibi yapabilmesi dinamik modelin doğruluğuna oldukça bağlıdır. Modellenemeyen aerodinamik etkiler ya da değişken yükler MPC perfomansını olumsuz etkiler. Esnek uçuş koşullarına hızlı uyum sağlayabilen bir MPC yöntemi, uyarlanabilen kontrolcülerin sisteme eklenmesiyle sağlanabilir [7].

Dört pervaneli dinamik modelini kullanan diğer bazı kontrol yöntemleri doğrusal karesel düzenleyici (*Linear Quadratic Regulator*, LQR) [8], geri-adımlama (*backstepping controller*) [9] ve kayan kipli kontrol (*sliding mode control*) [10] olarak sıralanabilir.

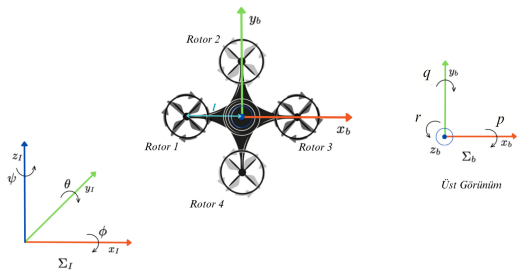
Dinamik model üzerinden kontrol sağlayan bu yöntemlerden farklı olarak derin sinir ağları (Deep Neural Networks, DNN) herhangi bir dinamik modele ihtiyaç duymadan doğrusal olmayan karmaşık sistemlerin yörünge kontrolünü kararlı bir şekilde sağlayabilir [11]. Derin belirgin yöntem eğilimi (*Deep Deterministic Policy Gradient*, DDPG) modelden bağımsız, gerçek zamanlı bir pekiştirmeli öğrenme yöntemidir. DDPG yönteminde aktör-eleştirmen pekiştirmeli öğrenme etmenleri uzun vadeli ortalama toplam ödülü en büyükleyecek politikayı arar [12]. Gürültüsüz ve bozucusuz simülasyon ortamında DDPG algoritması dört pervaneli kontrolü kabul edilebilir bir kalıcı hal hatası ile başarılı bir şekilde yapılmıştır [13]. DNN tabanlı kontrolcülerin sistem dinamikleri ile desteklenmesi hata performanslarını iyileştirmiştir [14].

Bu çalışmada dört pervanelilerin hareket ve yükseklik kontrolü için pekiştirmeli öğrenme tabanlı bir DDPG yöntemi geliştirilmiş, performansı çeşitli simülasyonlarla test edilmiştir. Bu makalede, dört pervaneli sistem dinamiği Bölüm 2'de tanıtılmış, önerilen pekiştirmeli öğrenme modeli Bölüm 3'te açıklanmış, sistem simülasyonları Bölüm 4'te tartışılmıştır. Bölüm 5 bulguları özetlemekte ve gelecek çalışmalar üzerine öneriler sunmaktadır.

## 2. Sistem Dinamikleri

Dört pervaneli uçuş dinamiği Newton'un ikinci hareket kanunu ve Newton-Euler denklemleriyle ifade edilebilir. Bunun için öncelikle sabit ve hareketli eksen takımların tanımlanması gerekir. Bir $\Sigma$ eksen takımı, origin noktasının koordinatları $o$ ve $(x, y, z)$ eksenlerin yönleri verilerek tanımlanabilir. Böylece, sabit yer eksen takımı $\Sigma_I = (o_I, x_I, y_I, z_I)$ ve hareketli hava aracının gövde eksen takımı $\Sigma_b = (o_b, x_b, y_b, z_b)$ ile ifade edilebilir. Bir eksen bizden uzaklaşır şekilde alındığında, saat yönündeki dönme artı, tersi yöndeki dönme eksi olarak kabul edilir [3].

$\Sigma_I$ eksen takımında $x_I$ pusula kuzeyini, $y_I$ doğuyu, ve $z_I$ yerçekiminin tersi yönü gösterir. Bir dört pervanelinin $\Sigma_b$ eksen takımında, $z_b$ ekseni $z_I$ ekseni ile aynı yönde iken $(x_b, y_b)$ eksenleri sağ el kuralına uygun olarak rotor eksenleri üzerindedir (Şekil 1). Her bir rotor gövdeye $l$ $(m)$ uzaklıktadır. Dört pervanelinin havada dengede kalabilmesi için rotor 1 ve 3 saat yönünde aynı hızla dönerken rotor 2 ve 4 aynı hızla saat yönünün tersine dönmelidir.




Şekil 1: Dört pervaneli uçuş dinamiği.

Hareketli $\Sigma_b$ eksen takımında doğrusal hız vektörü $\gamma = [u\ v\ w]^T$ ve açısal hız vektörü $\Omega = [p\ q\ r]^T$ ile gösterilir. Sabit $\Sigma_I$ eksen takımında konum vektörü $\xi = [x\ y\ z]^T$ iken yönelim vektörü $\eta = [\phi\ \theta\ \psi]^T$ Euler açılarıdır. Burada, $\phi$ yalpa (*roll*) açısı $x_I$ eksenine göre dönmeyi, $\theta$ yunuslama (*pitch*) açısı $y_I$ eksenine göre dönmeyi ve $\psi$ sapma (*yaw*) açısı $z_I$ eksenine göre dönmeyi ifade eder. Havacılık standartlarına göre $zyx$ dönme sıralaması esas alınır. $\Sigma_b$ eksen takımında $\gamma$ doğrusal hızıyla hareket eden bir hava aracı $\Sigma_I$ eksen takımında $\dot{\xi} = R\gamma$ doğrusal hızındadır.

$$R = \begin{bmatrix} c_\psi c_\theta & c_\psi s_\theta s_\phi - s_\psi c_\phi & c_\psi s_\theta c_\phi + s_\psi s_\phi \\ s_\psi c_\theta & s_\psi s_\theta s_\phi + c_\psi c_\phi & s_\psi s_\theta c_\phi - c_\psi s_\phi \\ -s_\theta & c_\theta s_\phi & c_\theta c_\phi \end{bmatrix} \quad (1)$$

Denklem (1), bu koordinat dönüşümünü sağlayan $R$ rotasyon matrisini tanımlar. Burada $c_\alpha = cos(\alpha)$ ve $s_\alpha = sin(\alpha)$ demektir. $R$ rotasyon matrisi küçük $(\phi, \theta, \psi)$ dönüşleri için $c_\alpha = 1$ ve $s_\alpha = \alpha$ varsayımıyla güncellenebilir. Böylece bir dört pervanelinin durumu, üç eksenel $[x_b, y_b, z_b]^T$ ve üç açısal $[\phi, \theta, \psi]^T$ konum olmak üzere altı serbestlik derecesiyle ifade edilebilir.

$\omega_i$ $(rad/sn)$ açısal hızla dönen bir $i$ rotoru $\Sigma_b$ eksen takımında Denklem (2) ile verilen $F_i$ $(N)$ itme kuvvetini üretir. Burada $k_f$ kuvvet sabitidir. Tüm rotorların saat yönünde (ve sa-

atin tersi yönünde) dönmesiyle oluşan net $F_{itme}$ kuvveti dört pervanelinin kalkışını (ve inişini) sağlar (Denklem (3)).

$$F_i = k_f \omega_i^2 \tag{2}$$

$$F_{itme} = F_1 + F_2 + F_3 + F_4 \tag{3}$$

$i$ rotoru dönerken hava sürtünmesinden kaynaklı dönme ekseninin tersine oluşan moment $\tau_i$, rotorların itme kuvvetlerinden kaynaklı $x_I$ eksenindeki $\tau_\phi$, $y_I$ eksenindeki $\tau_\theta$ ve $z_I$ eksenindeki $\tau_\psi$ momentleri Denklem (4)–(7) ile verilmiştir. Burada $k_m$ moment sabitidir [15].

$$\tau_i = k_m \omega_i^2 \tag{4}$$

$$\tau_\phi = l[(F_1 + F_4) - (F_2 + F_3)] \tag{5}$$

$$\tau_\theta = l[(F_3 + F_4) - (F_1 + F_2)] \tag{6}$$

$$\tau_\psi = \tau_1 - \tau_2 + \tau_3 - \tau_4 \tag{7}$$

Böylece $\Sigma_b$ eksen takımındaki kuvvet vektörü $F_{gövde}$, moment vektörü $M_{gövde}$ ile verilir (Denklem (8)).

$$F_{gövde} = [0, 0, F_{itme}]^T, M_{gövde} = [\tau_\phi, \tau_\theta, \tau_\psi]^T \tag{8}$$

Benzer şekilde $\Sigma_b$ eksen takımında $\Omega$ açısal hızları verildiğinde $\Sigma_I$ eksen takımında Euler açılarının değişim hızları $\dot{\eta} = [\dot{\phi}, \dot{\theta}, \dot{\psi}]^T$ hesaplanabilir.

$$T = \begin{bmatrix} 1 & s_\phi t_\theta & c_\phi t_\theta \\ 0 & c_\phi & -s_\phi \\ 0 & \frac{s_\phi}{c_\theta} & \frac{c_\phi}{c_\theta} \end{bmatrix} \tag{9}$$

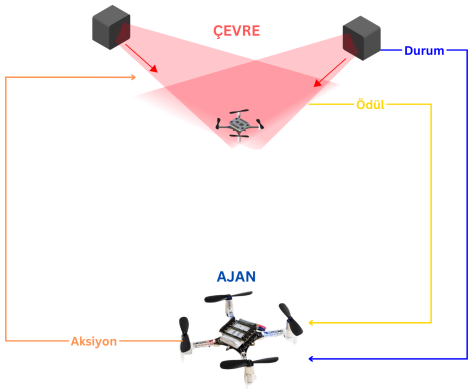
Denklem (9) ile tanımlanan $T$ dönüşüm matrisi $\dot{\eta} = T\Omega$ formülüyle açısal hız değişimini verir. Burada $t_\theta = tan(\theta)$ fonksiyonudur.

## 3. Pekiştirmeli Öğrenme Yöntemi

Pekiştirmeli öğrenme (*Reinforcement Learning*, RL) yönteminde kontrol kabiliyetine sahip bir etmen bir politika dahilinde aksiyon belirleyerek çevresiyle etkileşimde bulunur. Uygulanan aksiyonun çevreye uyumu, etmenin durumu ile hesaplanan bir ödül fonksiyonu aracılığıyla ölçülür (Şekil 2). Etmen, ödül fonksiyonunu en büyükleyecek şekilde aksiyonlarını belirleyen politikayı her adımda günceller. Böylece etmenin çevresiyle etkileşimini sağlayan en iyi politika belirlenmiş olur [16].

RL yöntemlerinin sürekli aksiyon uzayındaki karmaşık görevlerde karşılaştığı öğrenme zorlukları karar verici DNN alt sistemlerin eklenmesiyle aşılabilir. Bu yaklaşımla, değer tabanlı RL yöntemi (*Q-learning*) ve belirgin yöntem eğilimini (*Deterministic Policy Gradient*, DPG) birleştiren derin belirgin yöntem eğilimi (DDPG) metodu ortaya çıkmıştır [17].

DDPG yönteminde etmen, aktör (*actor*) ve eleştirmen (*critic*) ağı olmak üzere iki farklı DNN kullanır. $\pi$ politikasını kullanan bir aktör DNN, etmenin $s_t$ anlık durum girdisine en uygun $a_t$ aksiyonunu üretir ve $R(s_t, a_t)$ ödülünü kazanır. $V(s_t)$ durum değer fonksiyonu $s_t$ durumunda kazanılabilecek en büyük toplam ödülü ifade eder. Eleştirmen DNN ise etmenin $s_t$ anlık durumunu ve aktör DNN'den gelen $a_t$ aksiyonunu kullanarak bir $Q(s_t, a_t)$ durum-aksiyon değerini tahmin eder. Bu fonksiyon, $s_t$ durumunda $a_t$ aksiyonu uygulandığında kazanılabilecek en büyük toplam ödüldür.




Şekil 2: Pekiştirmeli öğrenme akış şeması.

Bellman denklemleri olarak bilinen Denklem (10) ile durum değerleri, Denklem (11) ile durum-aksiyon değerleri güncellenir. Burada $\gamma \in [0, 1]$ azaltma (*discount*) katsayısıdır. Belirlenen $V(s_t)$ değerleri her $s_t$ durumu için en iyi $a_t$ aksiyonunu verir ve aktör DNN'nin yeni politikasını oluşturur. DDPG öğrenme süreci sonunda sisteme en uygun aksiyon ve politika bulunmuş olur.

$$V(s_t) = \max_a \{R(s_t, a) + \gamma V(s_{t+1})\} \tag{10}$$

$$Q(s_t, a_t) = Q(s_t, a_t) + a(R(s_t, a_t) + \gamma \max Q(s_{t+1}, a_{t+1})) \tag{11}$$

Bölüm 3.1 dört pervanelilerin yükseklik ve yörünge kontrolü için geliştirdiğimiz DDPG yöntemini açıklamaktadır.

### 3.1. Derin Belirgin Yöntem Eğilimi

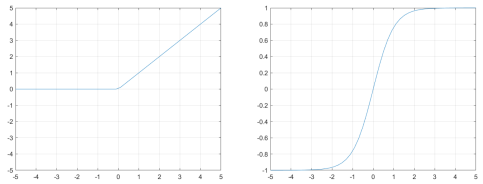
Uçuş kontrolü için önerdiğimiz DDPG yönteminde etmen bir dört pervanelidir. Etmenin kullandığı ileri beslemeli aktör ve eleştirmen DNN'leri, her biri 128 nöron içeren 4 katmandan oluşur (Şekil 3). Şekil 3(a)'da verilen aktör DNN, etmenin $\Sigma_l$ eksen takımındaki konum, hız, yönelim ve açısal hızından oluşan durum bilgisini girdi olarak alır. Böylece $s_t = (x, y, z, \dot{x}, \dot{y}, \dot{z}, \phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi})$ şeklinde tanımlanan 12 değişkenli bir vektördür. Aktör DNN'nin ürettiği aksiyon etmenin dört rotorunun $a_t = (\omega_1, \omega_2, \omega_3, \omega_4)$ hızlarıdır.


(a) (b)

Şekil 3: Etmen öğrenme ağları, (a) aktör DNN, (b) eleştirmen DNN.

Şekil 3(b)'de verilen eleştirmen DNN, etmenin $s_t$ durumuna ek olarak aktör DNN'nin hesapladığı $a_t$ rotor hızları ak-

siyonunu da ele alır. Böylece 16 girdili eleştirmen DNN, aktivasyon fonksiyonlarını uygulayarak bir $Q(s_t, a_t)$ ödül değeri hesaplar.



(a) (b)

Şekil 4: Aktivasyon fonksiyonları, (a) ReLu, (b) tanh.

Aktivasyon fonksiyonları bir DNN'nin doğrusal olmayan çıktılar üretmesini sağlar. Önerdiğimiz aktör ve eleştirmen DNN'lerin ilk üç katmanında kaybolan gradyanları düşük hesaplama yüküyle eleyebilmesi sebebiyle $f(x) = \max\{0, x\}$ şeklinde tanımlanan ReLU (*Rectified Linear Unit*) fonksiyonu kullanılmıştır [18]. DNN'lerin son katmanında, $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ aktivasyon fonksiyonu ile pozitif ve negatif değerler doğrusal olmayan bir şekilde işlenir (Şekil 4).

# 4. Tartışma

Dört pervaneli bir etmenin uçuş kontrolünü sağlayan Bölüm 3.1'de verilen DDPG yöntemimizin simulasyon ortamındaki öğrenme süreci, yörünge takibi ve birim basamak irtifa cevapları bu bölümde incelenmiştir. Simülasyonlar Windows 11 işletim sisteminde MATLAB 2023a yazılımıyla yapılmıştır. Kullanılan donanım 16 GB RAM, AMD Ryzen 5 5600H CPU ve NVIDIA CUDA RTX 3060 GPU özelliklerine sahiptir. Karmaşık hesaplamalar içeren DDPG yönteminin MATLAB simülasyonu CPU üzerinde yapılırken, aktör ve eleştirmen DNN'lerin öğrenme sürecinde basit hesaplamalar yoğunlukla tekrarladığından GPU desteği alınmıştır.

DDPG yöntemi Crazyflie dört pervanelisi etmeni için Simulink üzerinde test edilmiştir [19]. Crazyflie sitemine ait fiziksel parametreler ve genel simülasyon parametreleri Tablo 1 ile verilmektedir.

Tablo 1: Simülasyon parametreleri.

| *Crazyflie Fiziksel Parametreleri* | |
|---|---|
| $m$, ağırlık | 0.022 kg |
| $l$, uzunluk | 0.042 m |
| $I_{xx}$, eylemsizlik momenti | 9.1914e-06 kg-m$^2$ |
| $I_{yy}$, eylemsizlik momenti | 9.1914e-06 kg-m$^2$ |
| $I_{zz}$, eylemsizlik momenti | 2.2800e-05 kg-m$^2$ |
| Motor eylemsizlik momenti | 2.97e-05 kg-m$^2$ |
| $k_f$, kuvvet sabiti | 2.75e-11 kg-m |
| $k_m$, moment sabiti | 1 kg-m$^2$ |
| *Genel Parametreler* | |
| $d_t$, zaman aralığı | 0.01 sn |
| $g$, yerçekimi ivmesi | -9.81 kg-m$^2$ |
| $\gamma$, DDPG azaltma katsayısı | 0.99 |
| $\alpha_e$, Eleştirmen öğrenme katsayısı | 0.01 |
| $\alpha_a$, Aktör öğrenme katsayısı | 0.001 |

### 4.1. DDPG Yönteminde Politika Öğrenmesi

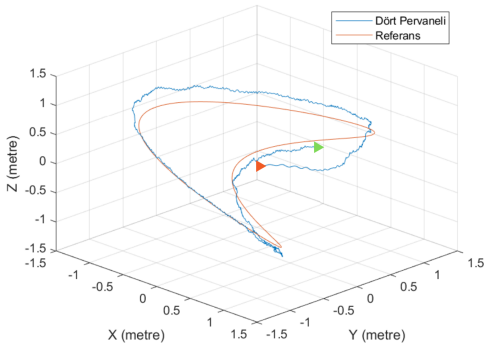
Crazyflie dört pervaneli simülasyonu için ilk adım DDPG yöntemine ait aktör ve eleştirmen DNN'lerin model parametrelerini öğrenmesidir. Bu amaçla, $5 \times 5 \times 5$ ($m^3$) boyutlarında merkezi $(0, 0, 0)$ koordinatlarıyla verilen sanal bir oda ele alınmıştır. Her öğrenme bölümünde (*train episode*) Crazyflie etmeni oda içerisinde rastgele bir $s_0 = (x, y, z, \dot{x}, \dot{y}, \dot{z}, \phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi})$ durumundan hareketine başlamıştır. Öğrenme süreci 100.000 bölüm devam etmiş ve her bölümde etmen 10 saniye uçuş gerçekleştirmiştir.

Etmen konum, hız ve Euler açılarına ait anlık durumunu değerlendirerek başlangıç konumundan odanın merkezine ulaşmayı hedefler. Bu doğrultuda, anlık duruma bağlı olarak her bir rotorun $a_t = (\omega_1, \omega_2, \omega_3, \omega_4)$ hızlarını aksiyon olarak öğrenmeye çalışır. Alınan $a_t$ aksiyonu odanın merkezine olan Öklit uzaklığı ve $\psi$ sapma açısınına bağlı olarak cezalandırılır. Yani, $a_t$ aksiyonuna ait $R(s_t, a_t)$ negatif bir ödüldür. Her öğrenme bölümünde etmen, cezayı en küçükleyecek rotor hızlarını aksiyon olarak öğrenir ve öğrenme sonunda en uygun $\pi$ politikasını üretir. Önerdiğimiz DDPG yönteminde etmen öğrenme sürecinin sonunda $\psi$ sapma açısını düşürmeyi ve odanın merkezine gitmeyi başarılı bir şekilde öğrenmiştir.

### 4.2. DDPG Yöntemi ile Yörünge Takibi

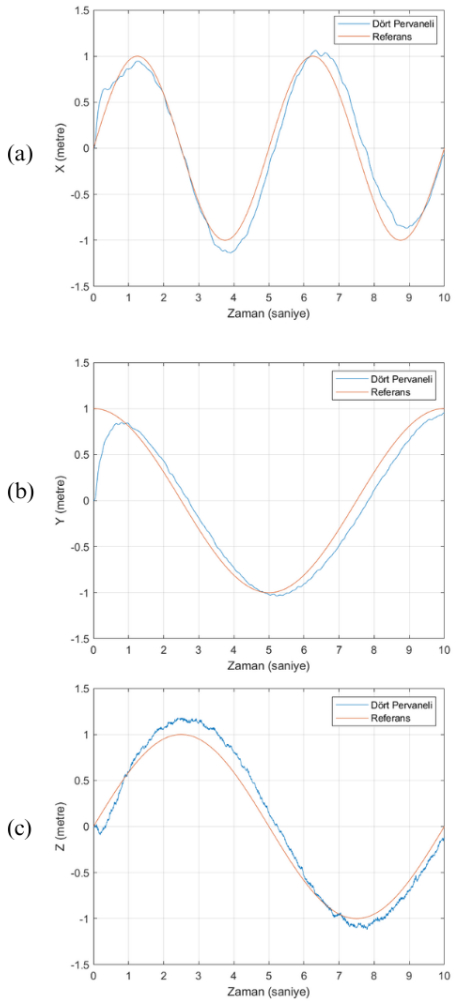
Bölüm 4.1'de anlatılan öğrenme sürecini tammalayan bir Crazyflie etmeninin yörünge takip performansını test etmek üzere $[x_t, y_t, z_t]^T = [s_{2t}, c_t, s_t]^T$ parametrik referans yörüngesi tanımlanmıştır. Burada, $t \in [0, 10]$ ($sn$) simülasyon anını ifade eder.

Şekil 5 referans yörüngeyi kırmızı eğri ile, etmen hareketini ise mavi yörünge ile göstermektedir. Etmen kırmızı üçgenle gösterilen ilk konumdan hareketine başlamış ve yeşil üçgenle gösterilen son konumuna gelinceye kadar referans yörüngeyi hafif sapmalarla yakından takip etmiştir.

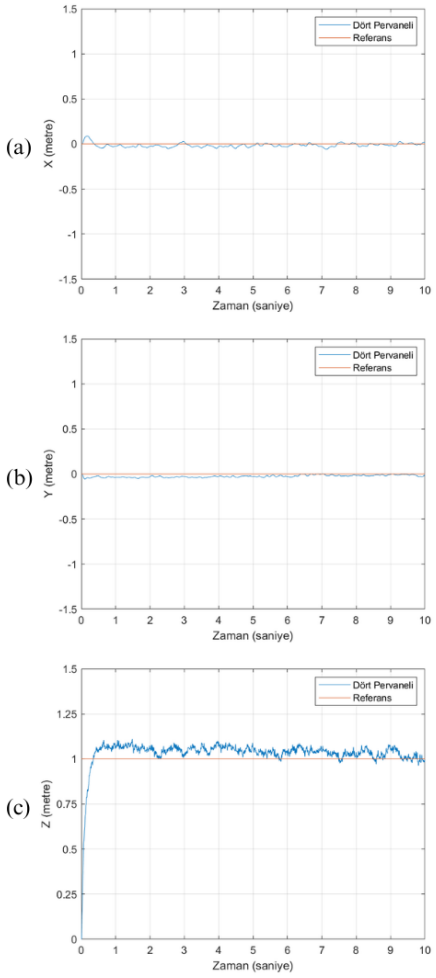



Şekil 5: 3B Yörünge takip simülasyon cevabı.

Etmenin $x$, $y$ ve $z$ eksenlerindeki detaylı yörünge takip performansı sırasıyla Şekil 6(a), 6(b) ve 6(c) ile verilmiştir. Oda içerisinde rastgele bir konumdan harekete başlayan etmen, referans yörüngeye yaklaşık 1 saniye içinde yakınsamış ve simülasyon süresi boyunca kabul edilebilir bir sapma ile referans yörüngeyi izleyebilmiştir.




Şekil 6: Yörünge takip simülasyonu, (a) $x$ ekseni, (b) $y$ ekseni, (c) $z$ ekseni cevabı.




Şekil 7: Birim basamak irtifa cevabı, (a) $x$ ekseni, (b) $y$ ekseni, (c) $z$ ekseni.

Sistem cevabında mavi yörünge incelendiğinde etmenin titreşimli hareketler yaptığı dikkat çekmektedir. Bu durum simülasyon ortam varsayımlarının hava sürtünmesi, pervane ataleti ve elastikliği gibi faktörleri içermemesinden kaynaklanabilir. Gerçek çevre koşullarında etmen sistem dinamiklerinin daha pürüzsüz bir yörünge üretmesi beklenmektedir.

### 4.3. DDPG Yöntemi ile Birim Basamak İrtifa Kontrolü

Birim basamak cevabı, bir kontrol sisteminin girişine birim basamak sinyal uygulandığında oluşan çıkış tepkisini gösteren, sistem performansını değerlendirmede sıklıkla kullanılan bir yaklaşımdır. Birim basamak testinde, $[x_i, y_i, z_i]^T = [0, 0, 0]$ ilk konumunda durmakta olan bir Crazyflie etmenine $[x_f, y_f, z_f]^T = [0, 0, 1]$ hedef konumu giriş sinyali olarak verilmiştir. Böylece, etmenin $x$ ve $y$ eksenlerinde sabit kalıp $z$ ekseninde bir metre yükselmesi istenmiştir. Etmenin $x$, $y$ ve $z$ eksenlerinde verdiği cevaplar sırayla Şekil 7(a), 7(b) ve 7(c) olarak gözlemlenmiştir. Etmen kendisinden beklenildiği üzere $x$ ve $y$ eksenlerinde pozisyonunu değiştirmezken $z$ ekseninde hedef irtifaya kademeli olarak ulaşmış ve bu konumunu sürdürmüştür.

Tablo 2 ile özetlenen kontrol performans ölçümleri, etmenin sergilediği başarılı birim basamak cevabını destekler niteliktedir. Etmen $T_r = 0.12$ saniye boyunca yükselmeye devam etmiş, hedef irtifayı en fazla $\%3$ aşmış, harekete başladığından $T_s = 3.8$ saniye sonra ise sabit irtifaya ulaşmıştır.

Tablo 2: Birim basamak cevabı performans değerleri.

| *Performans Ölçütü* | *Değer* |
|---|---|
| $T_r$, yükselme zamanı | 0.12 sn |
| $\%$ OS, yüzde aşım | $\%3$ |
| $T_s$, yerleşme zamanı | 3.8 sn |

Önerdiğimiz DDPG yönteminin gerek yörünge takibi gerekse birim basamak performansının literatürdeki benzer çalışmalar dikkate alındığında kabul edilebilir olduğu söylenebilir [12, 13, 16, 17]. Simülasyon ortamına dış faktörler ve gürültü gibi koşulların da dahil edilmesiyle geliştirilecek bir DDPG yöntemi dört pervaneli uçuş kontrolü için güçlü bir aday olacaktır.

## 5. Sonuçlar

Bu çalışmada dört pervaneli insansız hava araçlarının (İHA) irtifa kontrolü ve yörünge takibi için uçuş kontrolcü tasarımı ele alınmıştır. Bu amaçla dört pervaneli sistem dinamiği Newton-Euler yaklaşımı ve Newton'un hareket kanunları çerçevesinde modellenmiştir (Bölüm 2).

Dört pervaneli uçuş kontrolü için pekiştirmeli öğrenme (RL) ve belirgin yöntem eğilimi (DPG) yöntemlerini bir araya getiren bir derin belirgin yöntem eğilimi (DDPG) metodu geliştirilmiştir (Bölüm 3). DDPG yönteminde durum $(x, y, z)$ konumları ve hızları, $(\psi, \theta, \phi)$ Euler açıları ve hızları olmak üzere 12 değişkenden oluşmuştur. Aktör ve eleştirmen DNN'leri ile anlık duruma uygun rotor hızları aksiyon olarak belirlemiş, aksiyonun performansı bir ödül fonksiyonu ile değerlenmiştir. DDPG yöntemi ödülü en büyükleyecek aksiyonları öğrenerek bir politika belirlemiştir.

Belirlenen politika MATLAB ve Simulink ortamında Crazyflie dört pervaneli sistemi için simüle edilmiştir (Bölüm 4). Elde edilen sonuçlar, literatürdeki benzer çalışmalarla kıyaslandığında önerilen DDPG yönteminin bir dört pervanelinin irtifa ve yörünge kontrolünü yüksek başarımla sağlayabildiğini göstermektedir.

Bu bildiri kapsamında geliştirilen yöntemin uygulanabilirliğini ve literatürde sunulan benzer yöntemlerle yarışabileceğini göstermek amacıyla ideal uçuş şartlarında çalışılmıştır. Takip eden çalışmalarda önerilen yöntemin performansı farklı uçuş şartları ve bozucuların olması durumlarında sınanacaktır. Ayrıca kontrolör dayanıklılık açısından ele alınacak, detaylı kararlılık analizleri yapılacaktır.

DDPG yönteminde kullanılan aktör ve eleştirmen sinir ağlarının başarımını ve öğrenme hızını geliştirmek için fizik tabanlı sinir ağı (*Physics-Informed Neural Network*, PINN) kullanılacaktır. Ayrıca, bu sinir ağlarının performansını arttırmak için sabit ödül fonksiyonu yerine hedef öğrenmenin alt öğrenme adımlarının da hesaba katıldığı değişken ödül fonksiyonu modele dahil edilecektir.

## 6. Kaynakça


[1] S. Sun, G. Cioffi, C. De Visser, and D. Scaramuzza, "Autonomous Quadrotor Flight Despite Rotor Failure With Onboard Vision Sensors: Frames vs. Events", *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 580–587, 2021.

[2] M. Idrissi, M. Salami, and F. Annaz, "A Review of Quadrotor Unmanned Aerial Vehicles: Applications, Architectural Design and Control Algorithms," *J Intell Robot Syst*, vol. 104, no. 2, p. 22, 2022.

[3] P. Wang, Z. Man, Z. Cao, J. Zheng and Y. Zhao, "Dynamics modelling and linear control of quadcopter," 2016 Int. Conf. on Advanced Mechatronic Systems (ICAMechS), Melbourne, VIC, Australia, 2016, pp. 498-503.

[4] M. A. Khodja, M. Tadjine, M. S. Boucherit and M. Benzaoui, "Experimental dynamics identification and control of a quadcopter," 2017 6th Int. Conf. on Systems and Control (ICSC), Batna, Algeria, 2017, pp. 498-502.

[5] M. R. Kaplan, A. Eraslan, A. Beke, and T. Kumbasar, "Altitude and Position Control of Parrot Mambo Minidrone with PID and Fuzzy PID Controllers," in 2019 *11th Int. Conf. on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey: IEEE, Nov. 2019, pp. 785–789.

[6] F. Santoso, M. A. Garratt, S. G. Anavatti and I. Petersen, "Robust Hybrid Nonlinear Control Systems for the Dynamics of a Quadcopter Drone," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 8, pp. 3059-3071, 2020.

[7] D. Hanover, P. Foehn, S. Sun, E. Kaufmann, and D. Scaramuzza, "Performance, Precision, and Payloads: Adaptive Nonlinear MPC for Quadrotors," 2022.

[8] S. Günsel and Ş. N. Engin, "The Effects of PSO Parameters on an LQR Controlled Quadrotor System Gain," 2021.

[9] T. Madani and A. Benallegue, "Backstepping Control for a Quadrotor Helicopter," 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Beijing, China, 2006, pp. 3255-3260.

[10] H. Le Nhu Ngoc Thanh and S. K. Hong, "Quadcopter Robust Adaptive Second Order Sliding Mode Control Based on PID Sliding Surface," in IEEE Access, vol. 6, pp. 66850-66860, 2018.

[11] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. MIT Press, 1998.

[12] U. H. Ghouri, M. U. Zafar, S. Bari, H. Khan and M. U. Khan, "Attitude Control of Quad-copter using Deterministic Policy Gradient Algorithms (DPGA)," 2019 2nd Int. Conf. on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan, 2019, pp. 149-153.

[13] J. Hwangbo, I. Sa, R. Siegwart, and M. Hutter, "Control of a Quadrotor With Reinforcement Learning," *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 2096–2103, Oct. 2017.

[14] N. Bernini, M. Bessa, R. Delmas, A. Gold, E. Goubault, R. Pennec, S. Putot, and F. Sillion. "A few lessons learned in reinforcement learning for quadcopter attitude control." In Proceedings of the 24th Int. Conf. on Hybrid Systems: Computation and Control (HSCC '21). Association for Computing Machinery, 27, 1–11, 2021.

[15] T. Luukkonen, "Modelling and control of quadcopter." Independent research project in applied mathematics, Espoo, 22(22), 2011.

[16] I. Grondman, L. Busoniu, G. A. D. Lopes and R. Babuska, "A Survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy Gradients," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1291-1307, 2012.

[17] H. Liu, S. Suzuki, W. Wang, H. Liu, and Q. Wang, "Robust Control Strategy for Quadrotor Drone Using Reference Model-Based Deep Deterministic Policy Gradient," Drones, vol. 6, no. 9, p. 251, 2022.

[18] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," in Proceedings of the Fourteenth Int. Conf. on Artificial Intelligence and Statistics, pp. 315–323, 2011.

[19] Bitcraze. "Make your ideas fly." https://www.bitcraze.io (erişim, Agu. 01, 2023).