

Nümerik Gradyan İniş Yöntemi ve Sürekli Zaman Gradyan İniş Dinamiği Çözümlerinin Karşılaştırılması ve Lyapunov Kararlılığı

Comparision of Solutions of Numerical Gradient Descent Method and Continous Time Gradient Descent Dynamics and Lyapunov Stability

Nagihan Yağmur, Barış Baykant Alagöz

Bilgisayar Mühendisliği Bölümü,

İnönü Üniversitesi,

Malatya, Türkiye

nagihanyagmur95@gmail.com, baykant.alagoz@inonu.edu.tr

Özetçe—Gradyan iniş yöntemi makine öğrenmesi uygulamalarında yaygın olarak kullanılan bir optimizasyon tekniğidir. Bu teknik öğrenme hatasının azalan yönünde model parametrelerinin güncellemesini sağlanarak bir öğrenme dinamiği oluşturmaktadır. Bu çalışmada sürekli zaman gradyan iniş dinamiğinin Lyapunov kararlılık koşulları incelenmekte ve akıllı kontrol sistemi uygulamalarında gradyan iniş dinamiğinin uygulanabilmesi için ihtiyaç duyulan dayanıklı kararlılık koşulları incelenmektedir. Örnek bir uygulamada, De Jong's fonksiyonu türü hata fonksiyonu için gradyan iniş dinamiğinin ve Euler yöntemine dayanan nümerik çözümleri karşılaştırılmış ve kararlılığı konusu değerlendirilmiştir.

Anahtar Kelimeler — *gradyan iniş yöntemi; learning dynamics; Lyapunov kararlılığı.*

Abstract— Gradient descent dynamics is an optimization techniques that is widely used in machine learning applications. This technique updates model parameter in the direction of descending of learning error. In this study, Lyapunov stability of continuous time gradient descent dynamics is investigated and robust stability condition, which is needed for implementation of gradient descent dynamics in intelligent control system applications, is evaluated. In a illustrative example, for a De Jong's function type error function, solutions of continuous gradient descent dynamics and Euler method based numerical solutions are compared and stability concerns is discussed.

Keywords — *gradient descent method; learning dynamics; Lyapunov stability.*

I. GİRİŞ

Gradyan iniş yöntemi doğrusal olmayan programlamada (nonlinear programming) kullanılan temel bir optimizasyon tekniğidir[1]. Bu teknik basit çözümlemesi ve etkinliği nedeni ile makine öğrenmesi yöntemlerinde (yapay sinir ağları, derin

öğrenme vb.) nümerik optimizasyon için yaygın olarak kullanılmıştır [2, 4]. Genelde bir öğrenen sistem modeli için öğrenme performansını ifade eden bir öğrenme hata fonksiyonu, nümerik olarak gradyan iniş tekniği ile minimize edilerek öğrenme dinamiği oluşturulmaktadır [2]. Bir çok sinyal işleme, nesne tanıma uygulamasında öğrenme faaliyeti yapay sinir ağlarının sınıflayıcı olarak kullanılması ile gerçekleştirilmektedir. Bunun doğal bir sonucu olarak gradyan iniş yöntemini bu uygulamaların başarımını etkileyen önemli bir faktör haline dönüştürmüştür. Derin öğrenmenin gelişimi ile birlikte nümerik gradyan iniş tekniği optimal özellik çıkarma işlemi de üstlenerek sinyal ve görüntü işleme uygulamalarının en önemli bileşeni haline dönüşmüştür[3].

Gradyan iniş yönteminin kontrol sistemleri uygulamaları gerçekleştirilmiş olmasına rağmen [5, 6, 7, 8, 9], bu yöntem kontrol sistemleri pratiğinde sinyal ve görüntü işleme uygulamalarında görülen önem ve yaygın pratik kullanıma ulaşamamıştır. Bunun başlıca nedeni, gerçek zamanlı kontrol sistemi uygulamalarının, hızlı cevap ve dayanıklı kararlılığa ihtiyaç duymalarıdır. Gradyan iniş yöntemlerinin optimizasyon hızının iyileştirilmesi ve gradyan iniş çözümlerinin dayanıklı kararlılığının gösterilmesi, bu yöntemin gerçek zamanlı kontrol uygulamalarında daha yaygın kullanımına imkan sağlayacaktır. Bu çaba kontrol sistemlerinin öğrenebilen ve bulunduğu koşullara adapte olabilen akıllı kontrol sistemlerine dönüşebilmelerine katkı sağlayacaktır. Bu amaca dönük olarak, bu bildiride nümerik gradyan iniş yönteminin, sürekli zaman gradyan iniş dinamiğinin ayrık zaman Euler çözümü olduğu vurgulanmakta ve sürekli zaman gradyan iniş dinamiğinin dayanıklı kararlılık koşullu elde edilmektedir. De Jong's fonksiyonu türü kontrol hatası fonksiyonları için nümerik simülasyonlar yapılarak gradyan iniş dinamiğinin minimuma yakınsaması ve Euler

nümerik çözümlemesinin kararlılığı incelenecektir. Bu çalışmada elde edilen bulgular gradyan iniş dinamiğinin kontrol sistemlerinde daha etkin ve güvenilir bir şekilde uygulanmasına imkân sağlayabilir ve klasik kontrol uygulamalarının akıllı kontrol uygulamalarına dönüştürülmesine katkı sağlaması beklenebilir.

II. KISA LİTERATÜR ÖZETİ

A. Nümerik Gradyan İniş Yöntemi

Temel nümerik gradyan iniş yöntemi,

$$x_i[n+1] = x_i[n] - \gamma_i \frac{\partial F}{\partial x_i} \quad (1)$$

formülü ile ifade edilir. Burada $F(x)$ minimize edilmek istenen amaç fonksiyonudur. Parametreler $x_i \in R$, $x \in R^p$ vektörünün i . bileşenidir. Makine öğrenmesi uygulamalarında $F(x)$ hata fonksiyonu yaygın olarak öğrenme hatalarının karesel toplamı ile ifade edilir.

$$F(x) = \frac{1}{2} \sum_{j=1}^k e_j^2 \quad (2)$$

Burada, k eğitim kümesindeki veri sayısı, j . veri için anlık öğrenme hatası $e_j = y_j - y_e(x[n])$ ile ifade edilir [12].

Klasik nümerik gradyan iniş yöntemi her parametre güncellemesinde kullanılan veri miktarına göre veri analizinde üç farklı şekilde kullanılmıştır [12]: Tek bir güncelleme iterasyonunda eğitim setindeki bütün verileri kullanıldığı ($k \gg 1$) duruma yığın (batch) gradyan iniş yöntemi adı verilmiştir. Bu kullanım yüksek işlem maliyeti oluşturmaya karşın eğitim setinin öğrenilme performansını artırır. Her parametre güncellemesinde eğitim setinde sadece bir veri kullanıldığı durumda ($k = 1$) buna stokastik gradyan iniş yöntemi adı verilmiştir. İşlem maliyeti düşük olmasına rağmen öğrenme performansı diğerlerine göre düşük kalabilmektedir. Eğer parametre güncellemesi eğitim kümesinin birden fazla fakat tüm kümeden az alt kümeleri ile gerçekleşiyorsa buna mini yığın gradyan iniş yöntemi denilmiştir. Yığın gradyan inişten daha az işlem maliyeti ile stokastik gradyan inişten daha iyi öğrenme performansı sağlayabilmektedir [12].

Öğrenme algoritmalarının performansını artırabilmek için stokastik gradyan iniş yönteminin çeşitli varyantları geliştirilmiştir[12]: Moment tabanlı gradyan iniş yöntemleri[13], adaptif öğrenme katsayılı yöntemler [14, 15] vs. Gradyan iniş yönteminin kararlılık ve yakınsama koşulları için pek çok çalışma yapılmıştır [16, 17, 21].

B. Sürekli Zaman Gradyan İniş Dinamiği

Sürekli zaman gradyan iniş dinamiği

$$\frac{dx_i}{dt} = -\gamma_i \frac{\partial F}{\partial x_i} \quad (3)$$

diferansiyel denklemi ile ifade edilmiştir[18]. Adaptif kontrol uygulamalarında gradyan iniş işlemi MIT kuralı olarak adlandırılmıştır [8,19,20]. Bu denklem nümerik olarak Euler yöntemine göre çözmek için ileri fark denklemi [21],

$$\frac{dx_i}{dt} \cong \frac{x_i[n+1] - x_i[n]}{\Delta t} \quad (4)$$

kullanılırsa, denklem (1) ile ifade edilen temel nümerik gradyan iniş yöntemi formülasyonu elde edilir. Burada Δt parametresi benzetimin birim zaman adımı ya da nümerik çözümün örnekleme periyodu olarak ifade edilir.

III. GRADYAN İNİŞ DİNAMİĞİ KARARLILIK ANALİZİ

Bu çalışmada denklem (3) ile ifade edilen sürekli zaman gradyan iniş dinamiğinin kararlılık koşulları Lyapunov kararlılık kriteri [8, 10, 11, 21] yardımı ile elde edilmektedir.

Teorem 1 (Lyapunov kararlılık koşulu): Sürekli ve türevlenebilir bir $F(x): R^p \rightarrow R$ pozitif kesin (positive definite, $F > 0$) veya yarı-pozitif kesin (semi-positive definite, $F \geq 0$) reel değerli fonksiyonu için denklem (3) ile ifade edilen sürekli zaman gradyan iniş dinamiği $\forall \gamma_i > 0$ için her zaman (dayanıklı) kararlıdır.

İspat: Bir $F > 0$ veya $F \geq 0$ için Lyapunov enerji fonksiyonu,

$$V = F(x) \quad (5)$$

verilsin. $x_i \in R$ parametresi için Lyapunov enerji fonksiyonun Lyapunov kararlı olabilmesi için yeter koşul,

$$\frac{dV_i}{dt} \leq 0 \quad (6)$$

ile ifade edilir. Bu enerji fonksiyonun zamanla azalması veya sabit kalması durumunu ifade eder. Bu koşul uygulanırsa,

$$\frac{dV_i}{dt} = \frac{dF}{dt} = \frac{\partial F}{\partial x_i} \frac{dx_i}{dt} \leq 0 \quad (7)$$

yazılabilir. Burada $\frac{dx_i}{dt}$ için denklem (3) kullanılırsa,

$$\frac{dV_i}{dt} = \frac{\partial F}{\partial x_i} (-\gamma_i \frac{\partial F}{\partial x_i}) \leq 0 \quad (8)$$

Gradyan iniş dinamiğinin Lyapunov kararlı olabilmesi için yeter koşul,

$$-\gamma_i \left(\frac{\partial F}{\partial x_i} \right)^2 \leq 0 \quad (9)$$

elde edilir. Bu eşitsizliğin her zaman ve her x_i için geçerli olabilmesi için yeter koşulun $\forall \gamma_i > 0$ olması gerektiği kolaylıkla görülür.

Teorem 1, $\forall \gamma_i > 0$ koşulu altında denklem (3) ile ifade edilen sürekli zaman gradyan iniş dinamiğinin Lyapunov kararlı olacağını ve böylece $x_i(t)$ işaretinin zamanla $F(x)$ 'in azaltan yönde değişeceğini ifade eder. Bu teorem gradyan iniş yönteminin kontrol sistemi tasarımında uygulanabilirliği açısından önemli bir teorik dayanak sunmaktadır.

IV. ÖRNEK NÜMERİK ANALİZ

Bu bölümde hata fonksiyonlarını ifade edebilen, De Jong's fonksiyonu [22] türünde iki değişkenli,

$$F(x_1, x_2) = \sum_{i=1}^2 c_i x_i^2 \quad (10)$$

fonksiyonun sürekli zaman gradyan iniş dinamiğinin Euler yöntemi ile nümerik analizini yapalım. Burada c_1 ve c_2 katsayıları birer pozitif reel sayıdır. Sürekli zaman gradyan iniş dinamiğini bu fonksiyon için ifade etmek için,

$$\frac{\partial F(x_1, x_2)}{\partial x_1} = 2c_1 x_1 \quad (11)$$

$$\frac{\partial F(x_1, x_2)}{\partial x_2} = 2c_2 x_2 \quad (12)$$

elde edilir. Bu ifadeler denklem (3) kullanılırsa,

$$\frac{dx_1}{dt} = -\gamma_1 2c_1 x_1 \quad (13)$$

$$\frac{dx_2}{dt} = -\gamma_2 2c_2 x_2 \quad (14)$$

Elde edilen diferansiyel denklem sisteminin analitik çözümü,

$$x_1(t) = x_1(0)e^{-2c_1\gamma_1 t} \quad (15)$$

$$x_2(t) = x_2(0)e^{-2c_2\gamma_2 t} \quad (16)$$

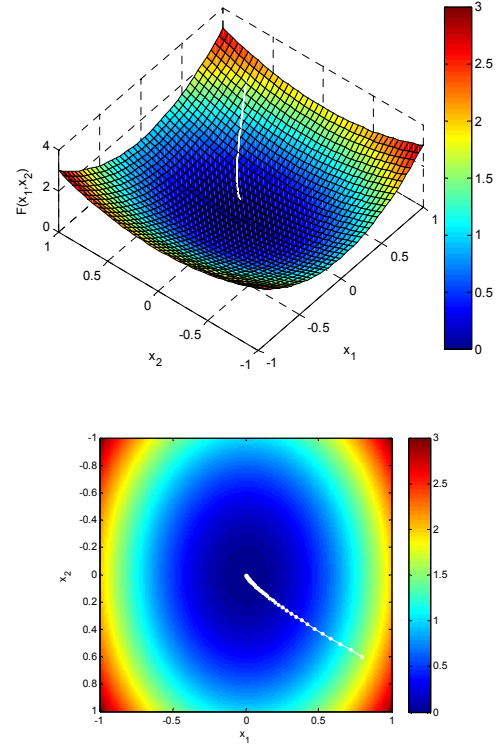
Elde edilen bu analitik çözümler nümerik çözümlerin doğruluğunu değerlendirmek için referans çözüm olarak değerlendirilecektir. Denklem (13) ve Denklem (14) Euler yöntemi ileri fark nümerik çözümü denklem (4) yardımı ile,

$$x_1[n+1] \cong x_1[n] - \Delta t \gamma_1 2c_1 x_1[n] \quad (17)$$

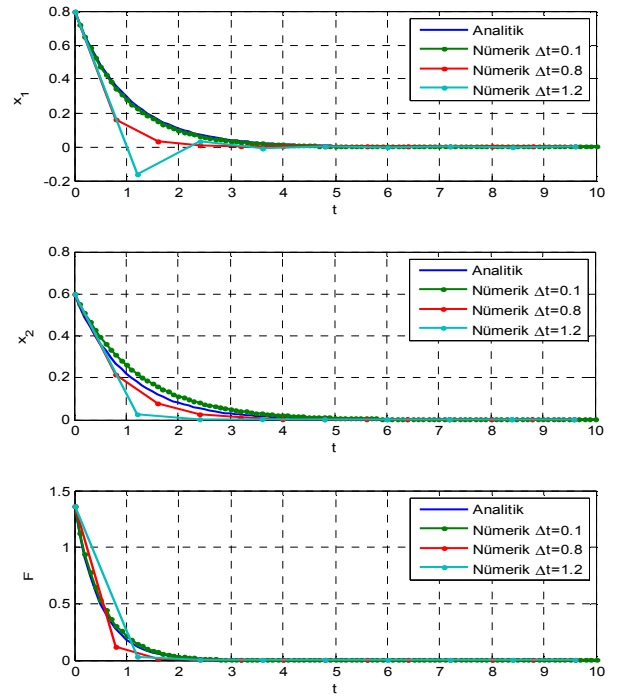
$$x_2[n+1] \cong x_2[n] - \Delta t \gamma_2 2c_2 x_2[n] \quad (18)$$

elde edilir. Burada $\eta_1 = \Delta t \gamma_1$ ve $\eta_2 = \Delta t \gamma_2$ çarpımları makine öğrenmesi uygulamalarında yaygın olarak öğrenme katsayısı olarak adlandırılmıştır. Şekil 1'de $c_1 = 1, c_2 = 2, \gamma_1 = 0.5, \gamma_2 = 0.2$ ve $\Delta t = 0.1$ için gradyan iniş dinamiğinin nümerik çözümü gösterilmiştir. Burada $x_1[n]$ ve $x_2[n]$ çözümlerinin başlangıç noktası $x_1[0] = 0.8$ ve $x_2[0] = -0.6$ noktasından minimuma giderken izlediği yörünge gösterilmiştir.

Şekil 2'de farklı Δt değerleri için gradyan iniş çözümleri görülmektedir. Δt değeri azaldıkça nümerik çözümün analitik çözüme yakınsadığı görülebilir.



Şekil 1. Nümerik gradyan iniş çözümünün yörüngesi



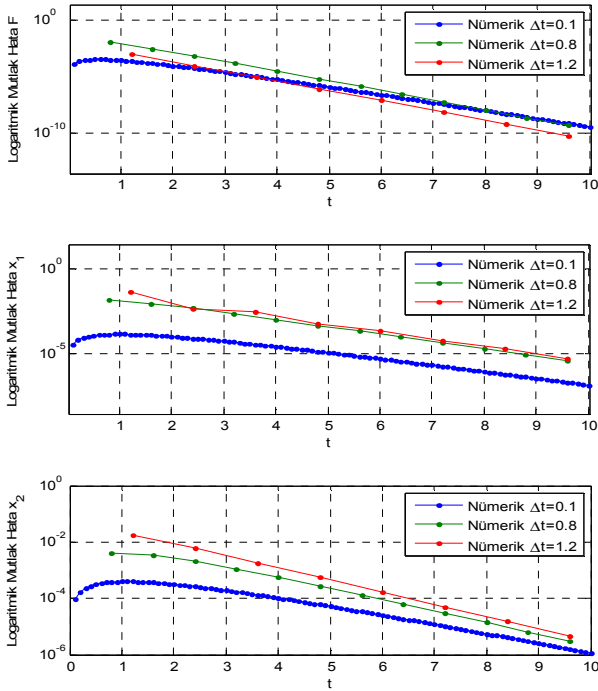
Şekil 2. Farklı örnekleme periyodları için nümerik çözüm ile analitik çözüm karşılaştırması

Tablo 1'de analitik çözüme göre nümerik çözümlerin ortalama mutlak hataları sunulmuştur. İkinci satırda mutlak hata artışı hızlı yakınsamanın sağladığı düşüşe karşın yakınsama hatalarının neden olabildiği artıştan kaynaklanmıştır.

TABLO I. NÜMERİK ÇÖZÜMÜN ORTALAMA MUTLAK HATA DEĞERLERİ

	F	x_1	x_2
$\Delta t = 0.1$	0.0055	0.0040	0.0118
$\Delta t = 0.8$	0.0162	0.0348	0.0117
$\Delta t = 1.2$	0.0120	0.0531	0.0260

Şekil 3’de Tablo 1’de görülen ortalama mutlak hataların anlık dağılımı görülmektedir. Düşük Δt için nümerik çözüm analitik çözüme başlangıçta daha iyi yakınsama sağlamayabilmesine rağmen $F(x_1, x_2)$ ’in sifıra yaklaştığı ilerleyen zamanda yüksek Δt yakınsama performansı iyileşebilmektedir. Bu hata dağılımları Tablo 1’de görünen mutlak hata değerlerine yol açan durumları açıklamaktadır.



Şekil. 3. Anlık mutlak hataların dağılımı

V. SONUÇLAR

Bu çalışmada sürekli zaman gradyan iniş dinamiğinin Lyapunov kararlılığı incelenmiş ve $\forall \gamma_i > 0$ için sürekli zaman gradyan iniş çözümünün her zaman Lyapunov kararlı olacağı görülmüştür. Örnek uygulamada nümerik gradyan iniş çözümünün De Jong's fonksiyonu varyantı için Euler çözümleri elde edilmiş ve sürekli zaman analitik çözümleri ile karşılaştırılmıştır. Sürekli zaman gradyan iniş dinamiği kararlı olmasına rağmen nümerik Euler çözümünün yüksek örneklem periyotlarında çözümünün mutlak hata performansının düştüğü ve kararsızlığa gidebildiği görülmüştür. Bu bulgular gradyan iniş tabanlı öğrenme dinamiğinin performansının ve kararlılığının nümerik

çözümleme tekniğinin yakınsama ve kararlılık performansına bağlılığını da gösterilmiştir.

Kaynaklar

- [1] Bertsekas, D. P., *Nonlinear Programming*, Athena Scientific, Belmont, 1999.
- [2] Amari, S. I., "Backpropagation and stochastic gradient descent method.", *Neurocomputing*, 5(4-5):185-196, 1993.
- [3] Amari, S. I. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5), 185-196.
- [4] Yin, P., Zhang, S., Lyu, J., Osher, S., Qi, Y., & Xin, J. (2019). Blended coarse gradient descent for full quantization of deep neural networks. *Research in the Mathematical Sciences*, 6(1), 14.
- [5] Dheeraj, K., Jacob, J., & Nandakumar, M. P. (2019). Direct Adaptive Neural Control Design for a class of Nonlinear Multi Input Multi Output Systems. *IEEE Access*.
- [6] Shimizu, K., Ito, S., & Suzuki, S. (2005). Tracking control of general nonlinear systems by a direct gradient descent method. *Nonlinear Dynamics and Systems Theory*, 5(1), 91-105.
- [7] Alagoz, B. B., Tepljakov, A., Kavuran, G., & Alisoy, H. (2018, September). Adaptive Control of Nonlinear TRMS Model by Using Gradient Descent Optimizers. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1-6). IEEE.
- [8] Alagoz, B. B., Kavuran, G., Ates, A., & Yeroglu, C. (2017). Reference-shaping adaptive control by using gradient descent optimizers. *PloS one*, 12(11), e0188527.
- [9] Bıdıklı, B. (2018). A Self-Tuning PID Control Method for Multi-Input-Multi-Output Nonlinear Systems. *Electrica*, 18(2), 218-226.
- [10] Khater, A. A., El-Nagar, A. M., El-Bardini, M., & El-Rabaie, N. M. (2018). Adaptive T-S fuzzy controller using reinforcement learning based on Lyapunov stability. *Journal of the Franklin Institute*, 355(14), 6390-6415.
- [11] Banakar, A. (2011). Lyapunov stability analysis of gradient descent-learning algorithm in network training. *ISRN Applied Mathematics*, 2011.
- [12] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- [13] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks : the official journal of the International Neural Network Society*, 12(1):145-151, 1999.
- [14] John Duchi, Elad Hazan, and Yoram Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12:2121-2159, 2011
- [15] Matthew D. Zeiler. ADADELTA: An Adaptive Learning Rate Method. *arXiv preprint arXiv:1212.5701*, 2012.
- [16] Fletcher, R., & Powell, M. J. (1963). A rapidly convergent descent method for minimization. *The computer journal*, 6(2), 163-168.
- [17] Hardt, M., Recht, B., & Singer, Y. (2015). Train faster, generalize better: Stability of stochastic gradient descent. *arXiv preprint arXiv:1509.01240*.
- [18] Baldi, P. (1995). Gradient descent learning algorithm overview: A general dynamical systems perspective. *IEEE Transactions on neural networks*, 6(1), 182-195.
- [19] Vinagre, B. M., Petráš, I., Podlubny, I., & Chen, Y. Q. (2002). Using fractional order adjustment rules and fractional order reference models in model-reference adaptive control. *Nonlinear Dynamics*, 29(1-4), 269-279.
- [20] Pankaj, S., Kumar, J. S., & Nema, R. K. (2011). Comparative analysis of MIT rule and Lyapunov rule in model reference adaptive control scheme. *Innovative Systems Design and Engineering*, 2(4), 154-162.
- [21] Wilson, Ashia. *Lyapunov Arguments in Optimization*. Diss. UC Berkeley, 2018.
- [22] Pohlheim, H. Geatbx examples examples of objective functions (2006). URL http://www.geatbx.com/download/GEATbx_ObjFunExpl_v37.pdf.