



# Kompleks düzlemde büyük-ölçekli regresyon: Bilgilendirici olmayan verileri çevrimiçi olarak sansürleyen CRLS algoritmalarının başarımlarını analiz

## Large-scale regression in the complex domain: Performance analysis of CRLS algorithms censoring noninformative data in an online manner.

Engin Cemal Mengüç<sup>1,\*</sup> 

<sup>1</sup> Kayseri Üniversitesi, Mühendislik, Mimarlık ve Tasarım Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 38280, Kayseri, Türkiye

### Öz

Büyük veri akışlarından anlamlı bilgilerin çıkarılması ve öğrenilmesi, toplumların yaşam kalitesinin artırılmasının, bilim ve mühendislik alanında yeni teknolojilerin geliştirilmesinin önünü açmaktadır. Öte yandan, sensör teknolojisindeki son atılımlar, hesaplama gücünün ve bilgisayar belleğinin artan kullanılabilirliği, verilerin sadece reel-değerli olmadığını artık büyük ölçekli kompleks-değerli veri kümeleriyle de başa çıkılması gerektiğini ortaya koymuştur. Bu amaç doğrultusunda, bu çalışmada, son zamanlarda önerilen çevrimiçi sansürleme (online censoring, OC) tabanlı kompleks-değerli özyinelemeli en küçük kareler (OC based complex-valued recursive least squares, OC-CRLS) ve OC tabanlı artırılmış CRLS (OC based augmented CRLS, OC-ACRLS) algoritmalarının başarımları ilk defa büyük ölçekli regresyon problemleri üzerinde detaylı olarak test edilmiş ve literatürde yer alan klasik versiyonları ile karşılaştırılmıştır. Benzetim çalışmaları, OC-CRLS ve OC-ACRLS algoritmalarının, OC mekanizmasının getirmiş olduğu avantajlardan dolayı kompleks düzlemde tanımlanmış olan büyük-ölçekli regresyon problemlerinde eğitim süresini ciddi anlamda kısalttığını ve test başarımlarını negatif yönde etkilemediğini göstermiştir. Bu da OC-CRLS ve OC-ACRLS algoritmalarının, kompleks düzlemde tanımlanabilen büyük veri akışı uygulamalarında etkin ve güçlü algoritmalar olduğunu kanıtlamıştır.

**Anahtar kelimeler:** Büyük veri, Regresyon, Çevrimiçi sansürleme, Kompleks-değerli veri, Özyinelemeli en küçük kareler.

### 1 Giriş

Verilerin bilgiye dönüştüğü, akıllıca kullanıldığında ve işlendiğinde ise insanlığa, bilim ve mühendislik dünyasına önemli katkılar sağlayacağı bir dijital veri tufanı çağında yaşamaktayız. Günümüzde bilimsel araçlar, tıbbi cihazlar, uydular, teleskoplar, sensörler, e-ticaret platformları, gözetim kameraları, cep telefonları, internet ortamı gibi birçok farklı kaynaktan üretilen eşi görülmemiş boyutlarda büyük veri akışları ile karşı karşıya kalmaktayız. Aslında büyük veri ve büyük veri akışları; yüksek boyut, karmaşıklık ve kullanılabilirlik oranı nedeniyle mevcut analitik yöntemlere meydan okuyan veri türlerini temsil etmektedir

### Abstract

Extracting and learning meaningful information from big data streams paves the way for improving the quality of life of societies and the development of new technologies in the field of science and engineering. On the other hand, recent advances in sensor technology, increased availability of computing power and computer memory reveal that data is not just real-valued, but large-scale complex-valued datasets must also be dealt with. For this purpose, for the first time in this study, the performances of the recently proposed online censoring (OC) based complex-valued recursive least squares (OC-CRLS) and OC based augmented CRLS (OC-ACRLS) algorithms are tested on large-scale regression problems and compared with those of their classical versions in the literature in detail. Simulation studies show that the OC-CRLS and OC-ACRLS algorithms significantly shorten the training time in large-scale regression problems defined in the complex domain without affecting testing performance in a negative way, due to the advantages of their OC mechanism. This proves that OC-CRLS and OC-ACRLS algorithms are effective and powerful algorithms in big data streaming applications that can be defined in the complex domain.

**Keywords:** Big data, Regression, Online censoring, Complex-valued data, Recursive least squares

[1]. Bu tür verilerden anlamlı bilgilerin türetilmesi ve öğrenilmesi, toplumların yaşam kalitesinin artırılmasının, bilim ve mühendislik alanında yeni teknolojilerin geliştirilmesinin önünü açmaktadır. Bu kapsamda, sayısal, matematiksel ve istatistiksel tekniklerin/yöntemlerin teorik alt yapılarla birleştirilerek yeni yöntemlerin geliştirilmesi günümüzde kaçınılmaz hale gelmiştir. Büyük veri büyük nimetlerle karşımıza çıkmaktadır fakat büyük zorluklara sahiptir [2-11]. Büyük verilerin doğası gereği yüksek hacim ve boyutu; geleneksel grup (batch) çıkarım yöntemlerinin oluşturulmasını imkânsız kılmaktadır [2-4]. Çünkü grup metodlar, en iyi sonucu üretebilmesi için ya verinin tamamına

\* Sorumlu yazar / Corresponding author, e-posta / e-mail: enginmenguc@kayseri.edu.tr (E.C. Mengüç)  
Geliş / Received: 15.01.2023 Kabul / Accepted: 25.02.2023 Yayınlanma / Published: 15.04.2023  
doi: 10.28948/ngumuh.1234303

ya da çok sayıda veriye ihtiyaç duyarlar. Bu yüzden de bu tür metotlar, ciddi depolama maliyeti gerektirir. Bilgi kaynaklarının çoğu; sürekli olarak akan, farklı çeşitlerde, miktarı ve istatistikleri zaman içinde ve elementler arasında hızla değişen büyük veri üretmektedir [2]. Dolayısıyla, büyük veri analitiğinin; gerçek-zamanlı kısıtlamalar altında geçmiş verileri tekrardan kullanmadan daha az veri kümelerinin zaman içinde sıralı olarak kullanıldığı çevrimiçi olarak gerçekleştirilmesi gerekmektedir [10,11]. Böylece, çevrimiçi yöntemlerle çok hızlı bir şekilde üretilen orta kaliteli bir çıktı, grup yöntemlerle oldukça yavaşça üretilen yüksek kaliteli bir çıktıdan daha yararlı olabilmektedir [5-8]. Bu nedenle, günümüzde büyük verinin çevrimiçi işlenmesi önemli hale gelmiştir. Öte yandan, büyük veri akışları yüksek derecede artık veri (redundancy) içerir; yani, yararlı bilgileri, orijinal boyutluklarına kıyasla çok daha düşük oranda olabilmektedir [11]. Büyük veri akışları, analiz için gerekli olan zaman ve kaynak ihtiyacı dikkate alındığında aslında kötü koşullandırılmış veri tipidir. Ayrıca verinin çeşitli kaynaklardan ve popülasyonlardan gelmiş olması, büyük verinin genellikle kirliliği (dirty) olarak tanımlanmasına yol açar [9]. Yani büyük veri akışları; doğal olarak gürültülü (bozuk) ölçümler/iletişim hataları içerir [2, 8, 9]. Diğer bir zorluk ise büyük veri akışlarında elzem olan görevlerin; yani, kestirim, tahmin ve regresyon görevlerinin yerine getirilmesidir [2, 8]. Eğer büyük veri akışları; seçilmiş bir veri alt kümesiyle kısıtlanabilirse, bahsi geçen problemler adaptif sinyal işleme (Adaptive signal processing, ASP) ile çözüme kavuşturulabilir.

Büyük verilerde, veri alt kümesiyle çalışabilmenin en kolay yolu; veri gruplarının yüksek derecede artıklık (redundancy) özelliğinden akılcıca faydalanmaktır [11]. Örneğin; çevrimiçi olarak önem arz eden veriyi önemsiz veriden ayırarak, büyük veri akışlarında veri işleme maliyeti ve veri boyutu ciddi bir şekilde azaltılabilir [11]. Bu kapsamda, sansürleme (censoring) işlemi son zamanlarda birçok farklı alanda doğal olarak popüler hale gelmiştir. Sansürlenmiş verilere dayanan grup kestirim (batch estimator) yöntemleri; ekonometri, biyometri, hayatta kalma analizi [12], doymuş ölçüm [13], spektrum algılama [14] dahil olmak üzere birçok mühendislik [15] görevinde kullanılmıştır. Son zamanlarda ise sansürleme tekniği; kaynak kısıtlı kablosuz sensör ağlarını [16, 17] kullanarak dağıtılmış parametre tahmini ve dinamik süreçler için veri seçiminde kullanılmış ve bu sayede sistem başarımı ile sistem maliyeti arasında bir denge sağlanmıştır. Bu çalışmalar, sansürlenmiş ölçümlerle elde edilen tahmin doğruluğunun, sansürsüz verilere dayalı olanlarla karşılaştırılabilir olduğunu doğrulamıştır. Bu yüzden, sansürleme işlemi; büyük veri uygulamalarında kesinlikle arzu edilen bir özellik olan veri işleme maliyetini düşürme potansiyeli sunmaktadır. Bu kapsamda, [11]'de yer alan çalışmada, büyük ölçekli doğrusal regresyonlarda büyük veri akışının boyutunu azaltmak ve veri işleme maliyetini düşürmek için çevrimiçi sansürleme (Online censoring, OC) stratejisi ilk defa kullanılmıştır. Bu çalışmada önerilen OC stratejisi altında yatan temel fikir; daha az bilgilendirici gözlemleri atmak için çevrimiçi aralık sansürlemesine dayanmaktadır. Son zamanlarda, büyük veri alanında

popüler hale gelen OC stratejisi [11,18]'de genellikle zamanla değişen parametrelerle dinamik süreçlerin izlenmesine ilişkin büyük veri sorunlarının, [19]'da enerji kaynaklarının sınırlı olduğu senaryolarda adaptif ağlar üzerinden dağıtılmış tahmin için enerji bütçesini düşürmenin ve [20]'de merkezi olmayan bir algılama sisteminde azaltılmış iletişim oranı elde etmenin üstesinden başarıyla gelmiştir. Bu çalışmalara ek olarak, OC stratejisi son zamanlarda merkezleştirilmiş büyük-ölçekli ağlarda [21], dağıtılmış kablosuz ağlarda [22], yapay sinir ağları eğitimi için bilgilendirici eğitim verisi çıkarmada [23], yapay sinir ağının eğitiminde [24], grafik sinyallerinin kestiriminde [25] ve doğrusal olmayan ağlar üzerinde adaptif filtreleme işlemlerinde [26], giyilebilir işleme cihazlarında akustik geri besleme yolunun kestiriminde [27], patolojik el tremorünün kestiriminde [28] kullanılmasıyla daha da popüler hale gelmiştir. Her ne kadar [11]'de tasarlanan OC stratejisi; büyük veri akışları için büyük-ölçekli doğrusal regresyon yöntemleri başlığı altında anılsa da, aslında literatürde klasik ASP algoritmaları olarak bilinen en küçük ortalama kare (Least mean square, LMS) ve özimizelemeli en küçük kareler (Recursive least squares, RLS) algoritmalarının yapısına yerleştirilmiştir (yani OC tabanlı LMS ve OC-tabanlı RLS algoritmaları önerilmiştir). Ayrıca, OC tabanlı algoritmalar, literatürde veri seçici algoritmalar olarak da bilinmektedir [29]. [2, 9, 11]'de OC stratejisi tasarlanırken, skaler akış gözlemleri klasik kestirim problemindeki gibi tanımlandığı için OC tabanlı LMS ve RLS algoritmaları doğal olarak ölçüm/iletişim hatalarını elemine etme özelliğine sahiptir. Bu da bu algoritmaların büyük veri akışlarında mevcut olan gürültü probleminin üstesinden geldiğini göstermektedir. Fakat önerilen bu algoritmalar, veriler aykırı değer içermediği sürece mantıklıdır; aksi durumda, bilgilendirici verilerin ayırt edilmesinde hatalı çalışmaktadır. Bu kapsamda yine aynı çalışmada [11], aykırı değere karşı gürbüzlük sağlayan modifiye edilmiş OC stratejisi tasarlanmış ve gürbüz OC tabanlı LMS ve RLS algoritmaları önerilmiştir. Bu sayede, büyük veri akışlarında olası aykırı değerlere karşı ek olarak önlem alınmıştır. Son zamanlarda, hem bilgilendirici olmayan hem de aykırı değerlerin sansürlendiği farklı sansürleme tabanlı (veri seçici) adaptif sinyal işleme algoritmaları [29]'da yer alan çalışmada sunulmuştur. Bu algoritmalar, veri seçici LMS, veri seçici RLS, veri seçici afin projeksiyon ve basitleştirilmiş veri seçici afin projeksiyon [29] algoritmaları olup, aslında temelinin [11]'deki çalışmaya dayandırıldığı görülmektedir. Bu tür OC tabanlı algoritmalar büyük veri akışının boyutunu azalttığı için büyük veri uygulamalarına özgü veri işleme ve depolama maliyetini önemli ölçüde azaltmıştır.

Sensör teknolojisindeki son atılımlar, hesaplama gücünün ve bilgisayar belleğinin artan kullanılabilirliği, verilerin sadece reel-değerli olmadığını artık büyük ölçekli kompleks-değerli veri kümeleriyle de başa çıkılması gerektiğini ortaya koymuştur [30]. Bu tür kompleks-değerli gerçek dünya verileri, fiziksel olarak önemli anlamlar taşıyan zengin bir yapıya sahiptir. Çünkü bu tür veriler, genlik ve faz bilgisini doğrudan içerir ve "reel ve imajiner" veya "faz ve genlik" olarak ifade edilebilirler. Kompleks-değerli veriler; ikinci-dereceden dairesel (second-order

circular) ve ikinci-dereceden dairesel olmayan (second-order noncircular) olmak üzere iki sınıfa ayrılır [31-38]. Literatürde, dairesel kompleks-değerli verileri işlemek için geleneksel kompleks-değerli ASP (Complex-valued ASP, CASP) algoritmaları geliştirilmiştir [31-38]. Bu kapsamda, [39]'da yer alan çalışmada, kompleks-değerli LMS (Complex-valued LMS, CLMS) algoritması önerilmiştir. Aslında bu önerilen algoritma, geleneksel LMS algoritmasının reel düzlemde kompleks bir uzantısıdır. Daha sonra ise, CLMS'nin farklı bir versiyonu olan en küçük faz-LMS (Least mean phase-least mean square, LMP-LMS) algoritması, faz ve genlik verilerinin aynı anda anlık olarak işlenmesi için [40]'da yer alan çalışmada önerilmiş olup, özellikle faz bilgisinin önemli olduğu birçok haberleşme uygulamasında kullanılmıştır. Ayrıca, diğer bir CASP algoritması, hızlı yakınsama kabiliyetine sahip olan ve yaygın olarak kullanılan kompleks-değerli RLS (Complex-valued RLS, CRLS) algoritmasıdır [41]. Bu bahsedilen CASP algoritmaları, ikinci-dereceden istatistiksel bilgi tabanlı algoritmalar; yani, istatistiksel bilgi olarak sadece kovaryans matrisi  $\mathbf{C}_{\mathbf{xx}} = E\{\mathbf{xx}^H\}$  bilgisini kullanırlar. Bu yüzden de bu tür algoritmalar sadece dairesel veriler için optimal çözümler sunarlar [31, 34, 42, 43], burada  $\mathbf{x}$ ,  $H$  ve  $E\{\cdot\}$  ifadeleri sırasıyla sıfır ortalamalı kompleks-değerli rastgele vektörü, eşlenik transpozunu ve beklenen değeri temsil etmektedir.

Son zamanlarda, artırılmış istatistikler CASP alanının daha da gelişmesini ve bu alanda yeni algoritmaların türetilmesini sağlamıştır. Çünkü artırılmış istatistikler; kovaryans matrisi  $\mathbf{C}_{\mathbf{xx}} = E\{\mathbf{xx}^H\}$ 'nin veriye ait ikinci-dereceden istatistiksel özellikleri tanımlamada artık yeterli olmadığını, özellikle dairesel olmayan verilerde ek olarak sözde-kovaryans  $\mathbf{P}_{\mathbf{xx}} = E\{\mathbf{xx}^T\}$  matrisinin de gerek olduğunu göstermiştir. Bu gereksinimleri yerine getirebilmek için artırılmış kompleks-değerli  $\mathbf{u} = [\mathbf{x}^T, \mathbf{x}^H]^T$  vektörü tanımlanarak  $\mathbf{x}$ 'e ait ikinci-hem kovaryans ve sözde kovaryans matrisleri, artırılmış kovaryans matrisi  $\mathbf{C}_{\mathbf{uu}} = E\{\mathbf{uu}^H\}$ 'nin içerisinde yerleştirilmiştir. Bu sayede hem dairesel hem de dairesel olmayan kompleks-değerli verilerin ikinci-dereceden istatistiksel özelliklerinin tamamı tek bir matris içerisinde değerlendirilmiştir [31-38]. Literatürde artırılmış istatistiklerin; biyomedikal, haberleşme, CASP, güç sistemleri gibi birçok farklı alanda sıklıkla kullanıldığı ve başarımı ciddi anlamda artırdığı gözlemlenmiştir [31-38, 42-53]. Artırılmış istatistiklerin ortaya çıkmasıyla, literatürde yer alan CASP algoritmaları; son zamanlarda kavramsal olarak kesin lineer (strictly linear, SL) ve geniş lineer (widely linear, WL) model tabanlı algoritmalar olarak ikiye ayrılmıştır. Örneğin; daha önce bahsettiğimiz CLMS, CRLS gibi kompleks düzlemde çalışan geleneksel algoritmalar, SL model tabanlı algoritmalar olarak adlandırılmaktadır. Literatürde, WL tabanlı algoritmalar ise ayrıca artırılmış kompleks-değerli algoritmalar olarak da bilinmektedir. Artırılmış istatistikler; WL model yapısını kullanan artırılmış kompleks-değerli en küçük ortalama kare (Augmented CLMS, ACLMS) [54] ve artırılmış kompleks-değerli özyinelemeli en küçük kareler (Augmented CRLS,

ACRLS) [55] algoritmalarının tasarımına yol açmıştır. Bu algoritmalar sayesinde hem dairesel hem de dairesel olmayan kompleks-değerli veriler kolaylıkla işlenebilmiştir. Hatta ACLMS ve ACRLS algoritmaları, dairesel olmayan veriler için CLMS ve CRLS algoritmalarından daha üstün bir başarımla sergilerken, dairesel veriler için CLMS ve CRLS ile aynı kararlı-durum hata cevabına daha yavaş yakınsama sergilemektedir [55, 56]. Bu yavaş yakınsamasının temel nedeni ise, bu algoritmaların WL yapılarından kaynaklı iki kat hesap yüküne sahip olmalarıdır [55-57]. Öte yandan, CRLS ve ACRLS algoritmaları, matris tersi alma işlemi [41, 55, 58] nedeniyle daha fazla hesaplama maliyeti pahasına daha hızlı yakınsama ve daha iyi kararlı durum hatası sunmaktadır. Genel olarak, CLMS, ACLMS, CRLS ve ACRLS algoritmaları; karmaşık-değerli verileri işlemek için çok etkili algoritmalar olsa da hem daha önce bahsedilen problemlerden hem de büyük veri işlemenin ve büyük sistem derecelerinin (filtre derecelerinin) getirmiş olduğu yüksek hesaplama maliyetlerinden muzdariptirler. Bu algoritmaların büyük veri akışlarında etkin bir şekilde kullanımını sağlamak için OC stratejisi dikkate alınarak yeniden tasarlanması gerekmektedir. Bu kapsamda, Mengüç ve arkadaşları [59]'daki çalışmalarında; CLMS, ACLMS, CRLS ve ACRLS algoritmalarının OC versiyonlarını türetmiş ve büyük-ölçekli kestirim ve tahmin problemleri üzerinde başarımlarını analiz etmiştir. Önerilen bu OC tabanlı algoritmalar [59], ağırlık vektörlerini sadece bilgilendirici veriler ile güncellediğinden kendilerinin klasik versiyonlarına kıyasla başarımı olumsuz yönde etkilemeden veri işleme maliyetini; yani, hesap yükünü ciddi anlamda düşürmüştür.

Bu çalışmada, son zamanlarda önerilen OC-CRLS ve OC-ACRLS algoritmalarının başarımları ilk defa regresyon problemleri üzerinde detaylı olarak test edilmiş ve literatürde yer alan ve klasik versiyonları olan CRLS ve ACRLS algoritmaları ile karşılaştırılmıştır. Benzetim çalışmaları; OC-CRLS ve OC-ACRLS algoritmalarının, OC mekanizmasının getirmiş olduğu avantajlardan dolayı kompleks düzlemde tanımlanmış olan büyük-ölçekli regresyon problemlerinde eğitim süresini ciddi anlamda kısalttığını ve aynı zamanda test başarımını negatif yönde etkilemediğini göstermiştir. Sonuç olarak, bu çalışma ile OC-CRLS ve OC-ACRLS algoritmalarının, kompleks düzlemde tanımlanabilen büyük veri akışı uygulamalarında etkin ve güçlü algoritmalar olduğu kanıtlanmıştır.

## 2 Materyal ve metod

### 2.1 Kesin lineer ve geniş lineer modeller

CASP perspektifinden, öncelikle SL model ile tanımlanan bir adaptif sistemin çıkışı aşağıdaki gibi tanımlanır:

$$y_k = \mathbf{x}_k^T \mathbf{h}_k \quad (1)$$

burada  $\mathbf{x}_k = [x_k, x_{k-1}, \dots, x_{k-N+1}] \in \mathbb{C}^{N \times 1}$  ve  $\mathbf{h}_k = [h_k, h_{k-1}, \dots, h_{k-N+1}] \in \mathbb{C}^{N \times 1}$  sırasıyla adaptif sistemin giriş vektörünü ve standart ağırlık vektörünü temsil eder.

Ayrıca  $N$  ifadesi ise sistemin derecesini (örneğin, çıkış sinyali bir sonlu dürtü cevaplı filtre (FIR) tarafından üretiliyorsa,  $N$  aynı zamanda filtre derecesidir) temsil eder. **Denklem (1)**'de yer alan SL model tanımı, kompleks-değerli dairesel verilerin işlenmesine olanak sağlar [31, 32].

Eğer sistem WL model kullanılarak tanımlanacak olursa, sistemin çıkışı  $y_k$ 'nın aşağıdaki gibi ifade edilmesi gerekir:

$$y_k = \mathbf{x}_k^T \mathbf{h}_k + \mathbf{x}_k^H \mathbf{g}_k \quad (2)$$

burada  $\mathbf{h}_k = [h_k, h_{k-1}, \dots, h_{k-N+1}] \in \mathbb{C}^{N \times 1}$  ve  $\mathbf{g}_k = [g_k, g_{k-1}, \dots, g_{k-N+1}] \in \mathbb{C}^{N \times 1}$  sırasıyla standart ve eşlenik ağırlık vektörlerini temsil eder. **Denklem (2)** artırılmış istatistik dikkate alınarak tanımlandığından tüm ikincideceden istatistiksel özellikleri yapısında barındırır ve hem dairesel hem de dairesel olmayan kompleks-değerli verilerin işlenmesine olanak sağlar [31, 32].

Bu iki model tanımının ardından hata sinyali,  $e_k = d_k - y_k = (e_k^r + j e_k^i) \in \mathbb{C}$  kullanılarak hesaplanır, burada  $j = \sqrt{-1}$ ,  $e_k^r$  ve  $e_k^i$  sırasıyla  $e_k$ 'ya ait reel ve imajiner kısımları,  $d_k$  ise ve beklenen (hedef) sinyali temsil eder.

## 2.2 Kompleks düzlemde çevrimiçi sansürleme

SL ve WL modeller dikkate alındığında, herhangi bir zaman indeksi  $k$  için beklenen skaler akış sinyali  $d_k \in \mathbb{C}$  aşağıdaki gibi modellenebilir:

$$d_k = \mathbf{u}_k^T \mathbf{w}_o + v_k, \quad k = 1, 2, \dots, K \quad (3)$$

burada  $\mathbf{w}_o$ ,  $\mathbf{u}_k$  ve  $v_k$  sırasıyla gerçek ağırlık vektörünü, giriş vektörünü ve giriş vektörüyle ilintisiz  $\sigma_v^2$  varyanslı dairesel simetrik sıfır ortalamalı beyaz kompleks Gauss ölçüm gürültüsünü temsil etmektedir. Ayrıca belirtmek gerekir ki; **Denklem (3)**'te SL model için  $\mathbf{w}_o = \mathbf{h}_o \in \mathbb{C}^{N \times 1}$  ve  $\mathbf{u}_k = \mathbf{x}_k \in \mathbb{C}^{N \times 1}$  olduğu, WL model için  $\mathbf{w}_o = [\mathbf{h}_o^T, \mathbf{g}_o^T]^T \in \mathbb{C}^{2N \times 1}$  ve  $\mathbf{u}_k = [\mathbf{x}_k^T, \mathbf{x}_k^H]^T \in \mathbb{C}^{2N \times 1}$  olduğu varsayılmaktadır. Burada  $\mathbf{h}_o$  gerçek standart ağırlık vektörü,  $\mathbf{g}_o$  gerçek eşlenik ağırlık vektörü ve  $\mathbf{x}_k = [x_k, \dots, x_{k-N+1}] \in \mathbb{C}^{N \times 1}$  ise giriş vektörüdür.

Sansürleme stratejisindeki ana amaç minimum kaynak gereksinimlerini kullanarak çevrimiçi olarak  $\mathbf{w}_o$ 'yı yani gerçek ağırlık vektörünü bulmaktır. Bu amaç doğrultusunda, ilk olarak FIR filtre yapısının çıkışının aşağıdaki gibi tanımlanması gerekir:

$$y_k = \mathbf{u}_k^T \mathbf{w}_k \quad (4)$$

burada  $\mathbf{w}_k$  ifadesi adaptif FIR filtrenin ağırlık vektörüdür. Hata sinyali ise aşağıdaki gibi ifade edilebilir:

$$e_k = d_k - y_k = d_k - \mathbf{u}_k^T \mathbf{w}_k \quad (5)$$

Böylece OC stratejisi herhangi bir CASP algoritması için aşağıdaki gibi kullanılabilir:

$$(z_k, c_k) := \begin{cases} (d_k, 0), & \text{eğer } \left| \frac{e_k}{\sigma_v} \right| \geq \tau \\ (0, 1), & \text{aksi durum} \end{cases} \quad (6)$$

burada  $c_k$  ve  $\tau_k > 0$  sırasıyla ikili sansürleme değişkenini ve  $k$  anındaki sansürleme eşik seviyesini temsil etmektedir.

**Denklem (6)**'dan gözlemlendiği üzere, eğer  $c_k = 1$  olursa, ilgilenilen akış verileri önce daha az bilgilendirici veri olarak etiketlenir ve ardından başka bir işlemde kullanılmadan sansürlenir. Eğer  $c_k = 0$  olur ise, ilgilenilen akış verileri önce en bilgilendirici veri olarak etiketlenir ve ardından güncelleme işlemine dahil edilir. **Denklem (6)**'da verilen OC stratejisi sayesinde her bir ölçüm anlık olarak "bilgilendirici (informative)" veya "bilgilendirici olmayan (noninformative)" ölçüm/verisi olarak sınıflandırılabilir. Sansürleme eşik seviyesi  $\tau$ 'nın seçimi, sansürleme oranını doğrudan etkilediğinden, bu eşik seviyesine; CASP algoritmalarının arzu edilen sansürleme oranı  $P_{ce}$ 'yi elde etmesi için bir değer atanması gerekir. Bu kapsamda  $\tau$  aşağıdaki denklem yardımıyla atanabilir [59]:

$$\tau = \sqrt{\ln \left( \frac{1}{1 - P_{ce}} \right)}. \quad (7)$$

burada  $P_{ce} = (K - p) / K$ 'dır ve  $K$  toplam veri sayısını temsil ederken,  $p$  bilgilendirici (sansürlenmemiş) veri sayısını temsil eder.

## 2.3 OC-ACRLS algoritması

Klasik CRLS ve ACRLS algoritmaları, bilindiği üzere deterministik maliyet fonksiyonu olan  $J_k = \sum_{n=0}^k \beta^{k-n} |e_n|^2$  fonksiyonu minimize edilerek türetilmektedirler. Aksine, OC-ACRLS algoritmasını türetmek için ise öncelikle bu deterministik maliyet fonksiyonunda yer alan  $|e_n|^2$  ifadesinin, **Denklem (6)**'da sunulan OC stratejisi dikkate alınarak aşağıdaki gibi tanımlanması gerekir [59]:

$$f_n := \begin{cases} |e_n|^2 - \tau^2 \sigma_v^2, & \text{eğer } |e_n| \geq \tau \sigma_v \\ 0, & \text{eğer } |e_n| < \tau \sigma_v \end{cases} \quad (8)$$

burada  $e_n = d_n - y_n = d_n - \mathbf{h}_{k-1}^H \mathbf{x}_n - \mathbf{g}_{k-1}^H \mathbf{x}_n^*$ , en son filtre ağırlık katsayıları kümesi kullanılarak hesaplanan zaman indeksi  $n$ 'deki hata sinyalidir.

Ardından ise **Denklem (8)** kullanılarak kesilmiş (truncated) maliyet fonksiyonu  $J_k^{(\tau)}$ , yeniden şu şekilde yazılır:

$$J_k^{(\tau)} := \sum_{n=0}^k \beta^{k-n} f_n \quad (9)$$

burada  $\beta \in (0,1]$  ifadesi üstel ağırlıklandırma (unutma) faktörünü temsil etmektedir.

ACRLS [55] algoritmasının türetilmesine benzer bir yol izlenerek maliyet fonksiyonu  $J_k^{(\tau)}$  ifadesi  $\mathbf{h}_{k-1}^*$  ve  $\mathbf{g}_{k-1}^*$ 'ye göre minimize edilir ise OC-ACRLS algoritmasının ağırlık vektörü güncelleme kuralları aşağıdaki gibi elde edilmiş olur [59]:

$$\text{Eğer } |e_k| \geq \tau\sigma_v, \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} + e_k^* \mathbf{k}_k \\ \mathbf{g}_k := \mathbf{g}_{k-1} + e_k^* \mathbf{k}_k^* \\ \mathbf{C}_k := \frac{1}{\beta} (\mathbf{C}_{k-1} - \mathbf{k}_k \mathbf{z}_k^H) \\ \mathbf{D}_k := \frac{1}{\beta} (\mathbf{D}_{k-1} - \mathbf{k}_k \mathbf{z}_k^T) \end{cases} \quad (10)$$

$$\text{Aksi durum, } \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} \\ \mathbf{g}_k := \mathbf{g}_{k-1} \\ \mathbf{C}_k := \mathbf{C}_{k-1} \\ \mathbf{D}_k := \mathbf{D}_{k-1} \end{cases}$$

burada  $\mathbf{z}_k = \mathbf{C}_{k-1} \mathbf{x}_k + \mathbf{D}_{k-1} \mathbf{x}_k^*$  ve  $e_k = d_k - \mathbf{h}_{k-1}^H \mathbf{x}_k - \mathbf{g}_{k-1}^H \mathbf{x}_k^*$ ,  $\mathbf{k}_k = \frac{\mathbf{z}_k}{\beta + 2\text{Re}\{\mathbf{z}_k^H \mathbf{x}_k\}}$ ,  $\mathbf{C}_k$  ve  $\mathbf{D}_k$  sırasıyla önsel kestirim

hatası, kazanç vektörü ve klasik kovaryans  $\mathbf{R}_{xx}$  ve sözde-kovaryans  $\mathbf{P}_{xx}$  matrislerinin terslerinin örnek kestirimleridir.

**Açıklama 1:** Giriş sinyalinin istatistiksel özelliklerinden bağımsız olan klasik ACRLS algoritması [55] dairesel ve dairesel olmayan kompleks-değerleri verilerin işlenmesinde üstün başarımla ortaya koymasına rağmen, WL modelinden kaynaklanan klasik  $\mathbf{R}_{xx}$  ve sözde-kovaryans  $\mathbf{P}_{xx}$  matrislerinin tersinin örnek kestirimleri nedeniyle yönetilemez bir hesaplama yüküne sahiptir. Denklem (10)'dan gözlemlendiği gibi; OC-ACRLS algoritması [59] klasik ACRLS'den farklı olarak, yapısındaki OC stratejisi sayesinde  $\mathbf{h}_k$ ,  $\mathbf{g}_k$ ,  $\mathbf{C}_k$  ve  $\mathbf{D}_k$ 'yi yalnızca  $|e_k| \geq \tau\sigma_v$  koşulu altında gelen bilgilendirici veriler için günceller. Bu önemli özelliği sayesinde, OC-ACRLS algoritması veri işleme maliyetini; yakınsama ve kararlı durum başarımlarını koruyarak, klasik ACRLS algoritmasına göre önemli ölçüde azaltır. Tablo 1'de OC-ACRLS algoritmasının sözde koduna ayrıca yer verilmiştir.

### 2.1 OC-CRLS algoritması

Denklem (9)'da yer alan maliyet fonksiyonu  $J_k^{(\tau)}$ 'nin, hata sinyali  $e_n = d_n - y_n = d_n - \mathbf{h}_{k-1}^H \mathbf{x}_n$  ile tanımlanması ve arından  $\mathbf{h}_{k-1}^*$ 'e göre minimize edilmesi sonucunda ise OC-CRLS [59] algoritmasına ait güncelleme aşağıdaki gibi elde edilir.

$$\text{Eğer } |e_k| \geq \tau\sigma_v, \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} + e_k^* \mathbf{k}_k \\ \mathbf{C}_k := \frac{1}{\beta} (\mathbf{C}_{k-1} - \mathbf{k}_k \mathbf{z}_k^H) \end{cases} \quad (11)$$

$$\text{Aksi durum, } \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} \\ \mathbf{C}_k := \mathbf{C}_{k-1} \end{cases}$$

burada  $\mathbf{z}_k = \mathbf{C}_{k-1} \mathbf{x}_k$ ,  $e_k = d_k - \mathbf{h}_{k-1}^H \mathbf{x}_k$  ve  $\mathbf{k}_k = \frac{\mathbf{z}_k}{\beta + \text{Re}\{\mathbf{z}_k^H \mathbf{x}_k\}}$  'dır.

**Açıklama 2:** OC-CRLS algoritmasının hesaplama karmaşıklığı, OC-ACRLS algoritmasının yaklaşık yarısı kadardır. CRLS'nin aksine OC-CRLS algoritması [59], veri işleme maliyetinde önemli tasarruf sağlar. Ayrıca, Tablo 2'de OC-CRLS algoritmasının sözde koduna yer verilmiştir.

**Tablo 1.** OC-ACRLS algoritmasının sözde kodu

Parametreler: $0 < \beta < 1$ ve $N$	
Başlangıç Koşulları: $\mathbf{h}_0 = \mathbf{g}_0 = \mathbf{0} + j\mathbf{0}$ , $\mathbf{C}_0 = \delta^{-1} \mathbf{I}$ ,	
$\mathbf{D}_0 = \mathbf{0}$ , $\hat{\sigma}_{v,0}^2 = 0$ , $P_{ce} = \frac{K-p}{K}$ ve $\tau = \sqrt{\ln\left(\frac{1}{1-P_{ce}}\right)}$	
Veri: $\{x_k, d_k\}_{k=1}^K$ .	
Döngü: $k = 1, 2, \dots, K$	
$y_k = \mathbf{h}_{k-1}^H \mathbf{x}_k - \mathbf{g}_{k-1}^H \mathbf{x}_k^*$	
$e_k = d_k - y_k = d_k - \mathbf{h}_{k-1}^H \mathbf{x}_k - \mathbf{g}_{k-1}^H \mathbf{x}_k^*$	
$\hat{\sigma}_{v,k}^2 = \rho \hat{\sigma}_{v,k-1}^2 + (1-\rho)  e_k ^2$	
$\mathbf{z}_k = \mathbf{C}_{k-1} \mathbf{x}_k + \mathbf{D}_{k-1} \mathbf{x}_k^*$	
$\mathbf{k}_k = \frac{\mathbf{z}_k}{\beta + 2\text{Re}\{\mathbf{z}_k^H \mathbf{x}_k\}}$	
Eğer $ e_k  \geq \tau\sigma_v$ , $\begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} + e_k^* \mathbf{k}_k \\ \mathbf{g}_k := \mathbf{g}_{k-1} + e_k^* \mathbf{k}_k^* \\ \mathbf{C}_k := \frac{1}{\beta} (\mathbf{C}_{k-1} - \mathbf{k}_k \mathbf{z}_k^H) \\ \mathbf{D}_k := \frac{1}{\beta} (\mathbf{D}_{k-1} - \mathbf{k}_k \mathbf{z}_k^T) \end{cases}$	
Aksi durum, $\begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} \\ \mathbf{g}_k := \mathbf{g}_{k-1} \\ \mathbf{C}_k := \mathbf{C}_{k-1} \\ \mathbf{D}_k := \mathbf{D}_{k-1} \end{cases}$	
Bitiş	

**Tablo 2.** OC-CRLS algoritmasının sözde kodu

Parametreler:  $0 < \beta < 1$  ve  $N$

Başlangıç Koşulları:  $\mathbf{h}_0 = \mathbf{0} + j\mathbf{0}$ ,  $\mathbf{C}_0 = \delta^{-1}\mathbf{I}$ ,

$$\hat{\sigma}_{v,0}^2 = 0, P_{ce} = \frac{K-p}{K} \text{ ve } \tau = \sqrt{\ln\left(\frac{1}{1-P_{ce}}\right)}$$

Veri:  $\{x_k, d_k\}_{k=1}^K$ .

Döngü:  $k = 1, 2, \dots, K$

$$y_k = \mathbf{h}_{k-1}^H \mathbf{x}_k$$

$$e_k = d_k - y_k = d_k - \mathbf{h}_{k-1}^H \mathbf{x}_k$$

$$\hat{\sigma}_{v,k}^2 = \rho \hat{\sigma}_{v,k-1}^2 + (1-\rho)|e_k|^2$$

$$\mathbf{z}_k = \mathbf{C}_{k-1} \mathbf{x}_k$$

$$\mathbf{k}_k = \frac{\mathbf{z}_k}{\beta + \text{Re}\{\mathbf{z}_k^H \mathbf{x}_k\}}$$

$$\text{Eğer } |e_k| \geq \tau \hat{\sigma}_v, \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} + e_k^* \mathbf{k}_k \\ \mathbf{C}_k := \frac{1}{\beta} (\mathbf{C}_{k-1} - \mathbf{k}_k \mathbf{z}_k^H) \end{cases}$$

$$\text{Aksi durum, } \begin{cases} \mathbf{h}_k := \mathbf{h}_{k-1} \\ \mathbf{C}_k := \mathbf{C}_{k-1} \end{cases}$$

Bitiş

### 3 Benzetim çalışmaları ve tartışma

Bu bölümde, OC-CRLS ve OC-ACRLS algoritmalarının öğrenme başarımı kompleks-değerli gerçek dünya regresyon problemleri üzerinde test edilmiştir. Bu kapsamda, iki adet gerçek veri seti kullanılmış. İlk veri kümesi, FIFA 19 Complete Player veri kümesidir [60]. İkinci veri kümesi ise büyük-ölçekli bir veri kümesi olan Block Feedback veri kümesidir [61]. Veri kümelerine ait bilgilere aşağıda yer verilmiştir.

*FIFA 19 veri kümesinde*, futbol oyuncularının tüm istatistikleri ve oyun özellikleri yer almaktadır. Toplamda 73 adet önceden işlenmiş özneliğe sahip, 14743 adet giriş-çıkış örneğinden oluşmaktadır. Öncelikli olarak giriş verileri, 1 ile -1 arasında olacak şekilde normalizasyon işlemine tabi tutulmuştur ve daha sonra aşağıdaki gibi kompleks-değerli forma dönüştürülmüştür:

$$\mathbf{x}_k = [1 + j1, x_{k,1} + jx_{k,2}, x_{k,3} + jx_{k,4}, \dots, x_{k,73} + j0]^T \quad (12)$$

burada kompleks-değerli her bir giriş verisi, ardışık reel-değerli çiftler ile eşleştirilmiş ve çift olmayan son giriş verisi ise  $j0$  değeri ile eşleştirilmiştir. Denklem (12)'de  $1 + j1$  değeri eşik değeri (bias) temsil etmektedir. Ayrıca belirtmek gerekir ki; verilerin kompleks düzlemde ifade edilmesi, reel düzleme kıyasla boyutu yarı yarıya azalmıştır. Bu da kompleks-düzlemin; artırılmış istatistik içermeyen OC-CRLS algoritması için önemli bir avantajı olarak ortaya çıkmaktadır.

*Block feedback veri kümesi* ise herhangi bir gönderinin kaç yorum alacağını tahmin etmek için kullanılan büyük-ölçekli bir veri kümesidir. Toplam 272 öznelikten (NaN öznelikleri hariç) oluşmakta olup, her bir özneliğin toplam örnek sayısı 52397'dir. Bir önceki veri kümesinde olduğu gibi, benzetimlerde giriş verileri, 1 ile -1 arasında olacak şekilde normalizasyon işlemine tabi tutulmuştur ve daha sonra aşağıdaki gibi kompleks-değerli forma dönüştürülmüştür:

$$\mathbf{x}_k = [1 + j1, x_{k,1} + jx_{k,2}, \dots, x_{k,271} + jx_{k,272}]^T \quad (13)$$

Benzetim çalışmalarında, algoritmaların regresyon başarımları ölçülürken ilgili veri kümelerinin %70'i eğitim aşamasında, %30'u ise test aşamasında kullanılmıştır. Eğitim aşamasındaki devir sayısı (epoch) 3000'e ayarlanmıştır. Ayrıca algoritmaların eğitim ve test başarımları; ortalama kare hata (Mean square error, MSE), regresyon katsayısı ve çalışma süreleri açısından incelenmiştir. Ayrıca, her iki veri kümesi için OC-CRLS ve OC-ACRLS algoritmalarının sansürleme oranı  $P_{ce} = \%50$  olarak ayarlanmıştır ve algoritmaların ait parametreler şu şekildedir.

FIFA 19 veri kümesi:

- CRLS, ACRLS, OC-CRLS ve OC-ACRLS için  $\lambda = 0.9995$ ,  $\rho = 0.55$ ,  $\mathbf{C}_0 = 10^2 \mathbf{I}$ ,  $\mathbf{D}_0 = \mathbf{0}$  ve  $\hat{\sigma}_{v,0}^2 = 0$  olarak seçilmiştir.

Block feedback veri kümesi:

- CRLS, ACRLS, OC-CRLS ve OC-ACRLS için  $\lambda = 0.9995$ ,  $\rho = 0.65$ ,  $\mathbf{C}_0 = 10^2 \mathbf{I}$ ,  $\mathbf{D}_0 = \mathbf{0}$  ve  $\hat{\sigma}_{v,0}^2 = 0$  olarak seçilmiştir.

FIFA 19 veri kümesi için Şekil 1, 2, 3 ve 4'te sırasıyla algoritmaların eğitim MSE (dB) eğrileri, eğitim regresyonları, test MSE (dB) eğrileri ve test regresyonları verilmiştir. Block feedback veri kümesi için ise Şekil 5, 6, 7 ve 8'de sırasıyla algoritmaların eğitim MSE (dB) eğrileri, eğitim regresyonları, test MSE (dB) eğrileri ve test regresyonları verilmiştir. Bu sonuçlara ek olarak ise Tablo 3 ve 4'te sırasıyla FIFA 19 ve Block feedback veri kümeleri için algoritmaların kestirilmiş sansürleme oranı, test MSE değerleri ve çalışma süreleri sunulmuştur. Elde edilen bu sonuçlara göre algoritmaların başarımları hakkında aşağıdaki yorumlar yapılmıştır.

Şekil 1'de FIFA 19 veri kümesi için algoritmaların eğitim MSE (dB) eğrileri incelendiğinde; önerilen OC-CRLS ve OC-ACRLS algoritmaları eğitim verilerinin sadece % 50'sini kullanarak, kendilerinin klasik eşlenikleri olan CRLS ve ACRLS algoritmalarına yakın başarımlar sergilemiştir. Burada, OC-CRLS ve OC-ACRLS algoritmaları yapılarında bulunan OC stratejisi sayesinde sadece bilgilendirici veriler ile eğitim sürecini gerçekleştirmiştir. Aynı zamanda, OC-ACRLS ve ACRLS algoritmaları artırılmış istatistikleri kullanan algoritmalar olduğundan dolayı OC-CRLS ve CRLS algoritmalarından daha üstün MSE (dB) başarımları elde etmiştir. Bu sonuçları daha da açığa kavuşturmak için Şekil 2'de aynı veri kümesi için algoritmaların eğitim

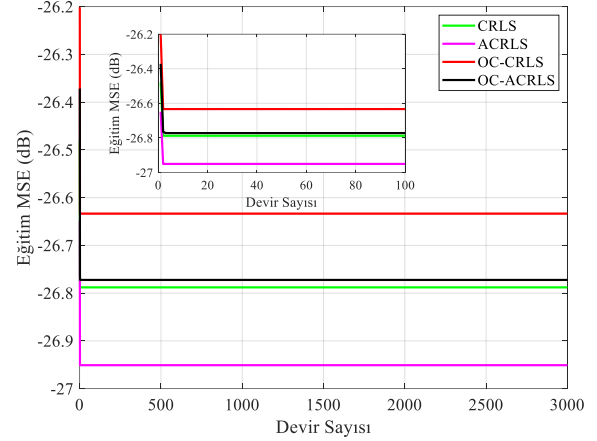
regresyon değerleri incelenmiştir. İncelenen bu sonuçlar, bütün algoritmaların, 0.89 değerine yakın korelasyon değerleri ürettiğini göstermiştir. Özellikle, OC-ACRLS algoritmasının diğer algoritmalarından bir miktar daha iyi korelasyon değeri ürettiği gözlemlenmiştir. Şekil 3'te ise eğitim süreci tamamlanan algoritmaların test MSE (dB) başarımları incelenmiştir. Şekil 3'ten görüldüğü üzere, OC-CRLS'nin test başarımları CRLS'den, OC-ACRLS'ninki ise ACRLS'den daha iyidir. Burada asıl vurgulanması gereken durum şudur: örneğin; önerilen OC-ACRLS algoritması eğitim aşamasında sadece bilgilendirici verileri kullanarak ve bu sayede eğitim süresini azaltarak, eğitim verilerinin % 100'ünü kullanan ACRLS algoritması ile neredeyse aynı eğitim MSE (dB) başarımları elde etmiştir. Ayrıca, Şekil 4'te test regresyon sonuçları incelendiğinde ise önerilen OC-CRLS ve OC-ACRLS algoritmaları en iyi korelasyon değerlerini elde etmiştir.

Şekil 5'te block feedback veri kümesi için algoritmaların eğitim MSE (dB) eğrileri incelendiğinde; artırılmış istatistik tabanlı kompleks-değerli algoritmaların tamamı kendilerinin klasik versiyonlarından daha üstün eğitim MSE (dB) başarımları göstermiştir. Şekil 5'ten gözlemlendiği üzere en iyi MSE (dB) başarımlarını hızlı yakınsama oranı ve düşük MSE (dB) değerleri ile OC-ACRLS ve ACRLS algoritmaları göstermiştir. Şekil 5'te verilen bu eğitim MSE (dB) sonuçlarını, aynı zamanda Şekil 6'da verilen regresyon sonuçları da desteklemektedir. Buradan görüleceği üzere, önerilen OC-ACRLS algoritmasının korelasyon değeri 1'e oldukça yakındır. Şekil 7'de ise eğitim sürecinden geçirilen bu algoritmaların test MSE (dB) başarımları incelenmiştir ve Şekil 6'daki sonuçlara paralel sonuçlar elde edilmiştir. Yine gözlemlendiği üzere önerilen OC-ACRLS algoritması eğitim aşamasında sadece bilgilendirici verileri kullanarak ve bu sayede eğitim süresini ciddi anlamda azaltarak, eğitim verilerinin % 100'ünü kullanan ACRLS algoritmasından oldukça üstün test MSE (dB) başarımları elde etmiştir. Ayrıca, Şekil 8'de test regresyon sonuçları incelendiğinde ise önerilen OC-ACRLS algoritması 1'e oldukça yakın korelasyon değeri elde etmiştir.

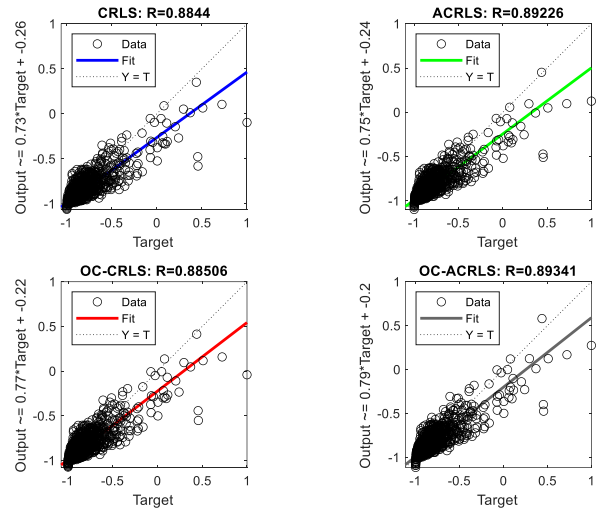
Tablo 3'ten gözlemleneceği üzere FIFA 19 veri kümesi için OC-CRLS ve OC-ACRLS algoritmaları için kestirilmiş sansürleme  $\hat{P}_{ce}$  değerleri, kullanıcı tarafından belirlenen gerçek sansürleme  $P_{ce}$  değerlerine oldukça yakındır. Ayrıca burada en iyi test MSE değerini OC-ACRLS algoritması elde etmiştir. Aynı zamanda, bu algoritmanın eğitim aşaması 405.6 saniye sürerken kendisinin klasik versiyonu olan ACRLS'in eğitim aşaması 768 saniye sürmüştür. Ayrıca, OC mekanizmasının OC-ACRLS algoritmasının eğitim süresini yaklaşık %50 oranında azalttığını burada belirtmekte fayda vardır. Bu sonuçlar, OC-ACRLS algoritmasının yüksek test MSE başarımlarından ve düşük eğitim süresinden dolayı büyük-ölçekli veri kümelerinin regresyonunda kolaylıkla kullanılabilirliğini göstermiştir.

Block feedback veri kümesi, FIFA veri kümesine göre daha büyük boyutlu olduğundan, Tablo 4'te, önerilen OC-ACRLS algoritmasının hem eğitim süresi hem de test MSE değeri açısından diğer algoritmalara göre daha tercih edilebilir olduğu görülmektedir. Örneğin, OC-ACRLS

algoritması çok düşük test MSE değerini yaklaşık 6860 saniyelik bir eğitimin ardından sağlarken kendisinin klasik versiyonu olan ACRLS algoritması daha yüksek MSE değerini yaklaşık 15633 saniyede sağlamıştır. Sonuç olarak; hem OC-CRLS hem de OC-ACRLS ciddi anlamda daha kısa eğitim süreleri sağlayarak, tatmin edici test MSE başarımları elde etmiştir.



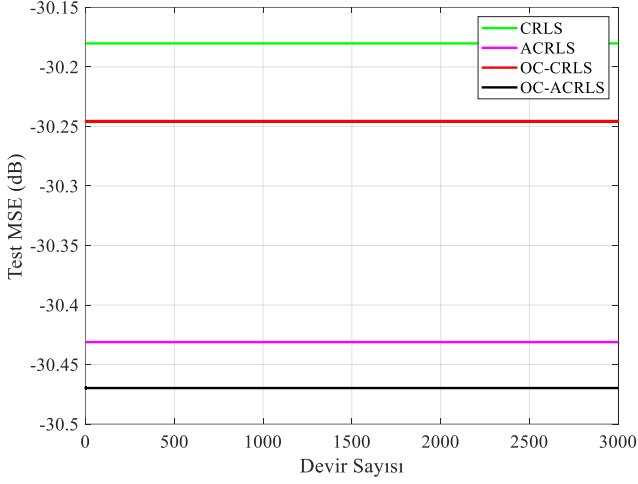
Şekil 1. FIFA 19 veri kümesi için algoritmaların eğitim MSE (dB) eğrileri



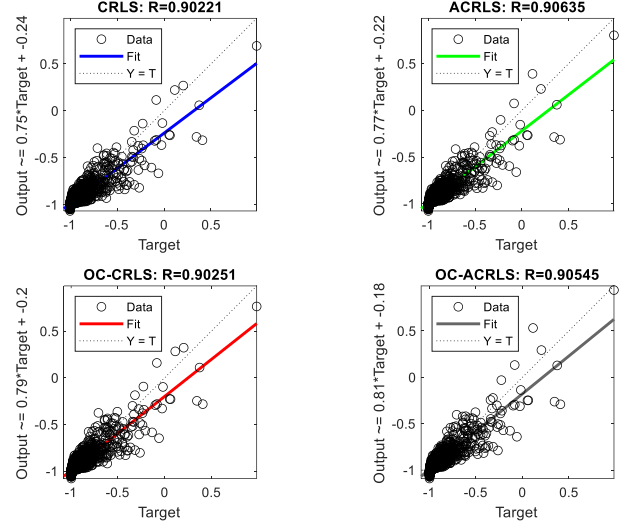
Şekil 2. FIFA 19 veri kümesi için algoritmaların eğitim regresyonları

Tablo 3. FIFA 19 veri kümesi için kestirilmiş sansürleme oranları, test mse değerleri ve çalışma süreleri,  $P_{ce} = \%50$

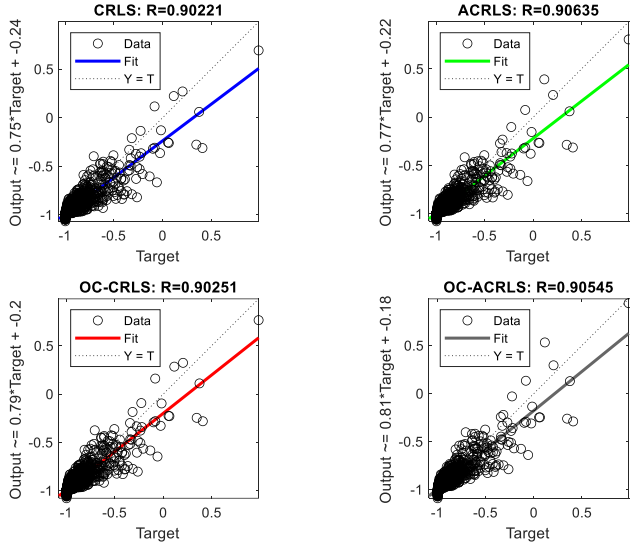
Algoritmalar	$\hat{P}_{ce}$	Test MSE	Çalışma Süresi (Saniye)
CRLS	-	9.5934e-04	379.5
OC-CRLS	% 56.82	9.4500e-04	210.9
ACRLS	-	9.0549e-04	768.9
OC-ACRLS	% 58.05	8.9748e-04	405.6



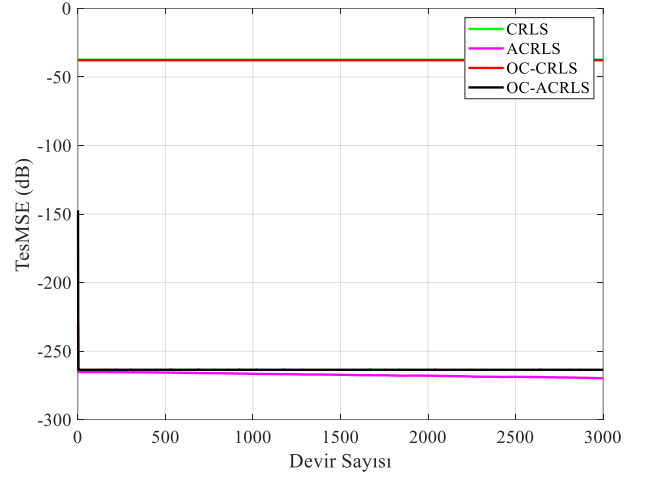
Şekil 3. FIFA 19 veri kümesi için algoritmaların test MSE (dB) eğrileri



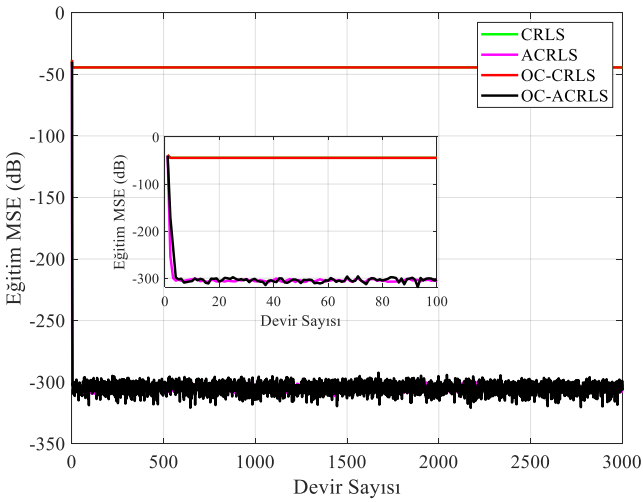
Şekil 6. Blok feedback veri kümesi için algoritmaların eğitim regresyonları



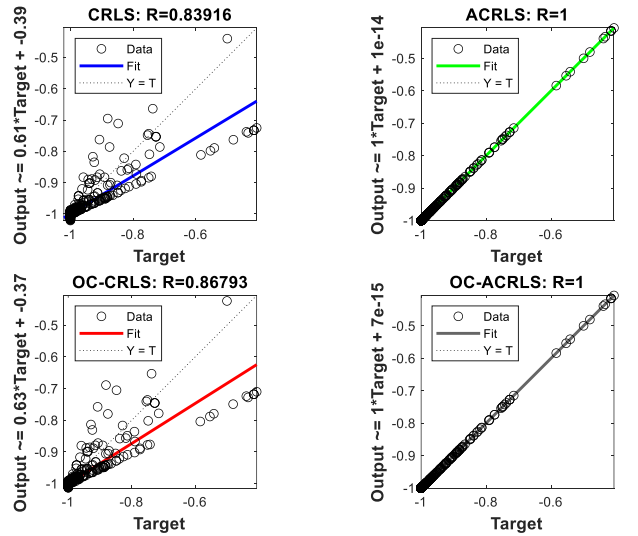
Şekil 4. FIFA 19 veri kümesi için algoritmaların test regresyonları



Şekil 7. Blok feedback veri kümesi için algoritmaların test MSE (dB) eğrileri



Şekil 5. Blok feedback veri kümesi için algoritmaların eğitim MSE (dB) eğrileri



Şekil 8. Blok feedback veri kümesi için algoritmaların test regresyonları



**Tablo 4.** Block feedback veri kümesi için kestirilmiş sansürleme oranları, test MSE değerleri ve çalışma süreleri,  $P_{ce} = \%50$

Algoritmalar	$\hat{P}_{ce}$	Test MSE	Çalışma Süresi (Saniye)
CRLS	-	1.8170e-04	7885.8
OC-CRLS	% 45.81	1.6707e-04	4901.7
ACRLS	-	1.0459e-27	15633
OC-ACRLS	% 64.15	4.3695e-27	6860.1

#### 4 Sonuçlar

Bu çalışmada, son zamanlarda önerilen OC-CRLS ve OC-ACRLS algoritmalarının başarımları ilk defa regresyon problemleri üzerinde detaylı olarak test edilmiş ve literatürde yer alan klasikleri (CRLS ve ACRLS) ile karşılaştırılmıştır. Benzetim çalışmaları, OC-CRLS ve OC-ACRLS algoritmalarının, OC mekanizmasının getirmiş olduğu avantajlardan dolayı kompleks düzlemde tanımlanmış olan büyük-ölçekli regresyon problemlerinde eğitim süresini yaklaşık % 50 oranda kısalttığını ve test başarımını negatif yönde etkilemediğini göstermiştir. Bu da OC-CRLS ve OC-ACRLS algoritmalarının, kompleks düzlemde tanımlanabilen büyük veri akışı uygulamalarında etkin ve güçlü algoritmalar olduğunu kanıtlamıştır.

#### Teşekkür

Bu çalışma kısmen Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından desteklenmiştir (Proje Numarası: 121E324).

#### Çıkar çatışması

Yazar çıkar çatışması olmadığını beyan etmektedir.

#### Benzerlik oranı (iThenticate): % 12

#### Kaynaklar

- [1] Z. Han, M. Hong and D. Wang, Signal Processing and Networking for Big Data Applications (1. Basım). United Kingdom: Cambridge Core, 2017.
- [2] K. Slavakis, G. B. Giannakis and G. Mateos, Modeling and optimization for big data analytics: (Statistical) learning tools for our era of data deluge. IEEE Signal Processing Magazine, 31 (5), 18-31, 2014.
- [3] T. Bengtsson, P. Bickel and B. Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. Probability and Statistics: Essays in Honor of David A. Freedman, 2, 316-334, 2008.
- [4] M. I. Jordan, On statistics, computation and scalability. Bernoulli, 19 (4), 1378-1390, 2013.
- [5] J. Mairal, F. Bach, J. Ponce and G. S.apiro, Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research, 11, 19-60, 2010.
- [6] M. Mardani, G. Mateos and G. B. Giannakis, Dynamic anomalography: Tracking network anomalies via sparsity and low rank. IEEE Journal on Selected Topics in Signal Processing, 7 (1), 50-66, 2013.

- [7] S. Theodoridis, K. Slavakis and I. Yamada, Adaptive learning in a world of projections: A unifying framework for linear and nonlinear classification and regression tasks. IEEE Signal Process, 28, 97-123, 2011.
- [8] J. Qiu, Q. Wu, Y. Xu and S. Feng, A survey of machine learning for big data processing. Eurasip Journal on Advances in Signal Processing, 1 (67), 1-16, 2016.
- [9] A. L'Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, Machine learning with big data: challenges and approaches. IEEE Access, 5, 7776-7797, 2016.
- [10] K. Slavakis, S. J. Kim, G. Mateos and G. B. Giannakis, Stochastic approximation vis-a-vis online learning for big data analytics. IEEE Signal Proc Mag., 31 (6), 124-129, 2014.
- [11] D. Berberidis, V. Kekatos and G. B. Giannakis, Online censoring for large-scale regressions with application to streaming big data. IEEE Trans. Signal Process., 64 (15), 3854-3867, 2016.
- [12] L. Evers and C. M. Messow, Sparse kernel methods for high-dimensional survival data. Bioinformat., 14 (2), 1632-1638, 2008.
- [13] J. Tobin, Estimation of relationships for limited dependent variables. Econometrica: J. Econometr. Soc., 26 (1), 24-36, 1958.
- [14] S. Maleki and G. Leus, Censored truncated sequential spectrum sensing for cognitive radio networks. IEEE J. Sel. Areas Commun., 31 (3), 364-378, 2013.
- [15] T. Amemiya, Tobit models: A survey. J. Econom., 24 (1), 3-61, 1984.
- [16] E. Msechu and G. B. Giannakis, Sensor-centric data reduction for estimation with WSNs via censoring and quantization. IEEE Trans. Signal Process., 60 (1), 400-414, 2012.
- [17] K. You, L. Xie and S. Song, Asymptotically optimal parameter estimation with scheduled measurements, IEEE Trans. Signal Process., 61 (14), 3521-3531, 2013.
- [18] D. Berberidis and G. B. Giannakis, Data sketching for large-scale Kalman filtering. IEEE Trans. Signal Process., 65(14), 3688-3701, 2017.
- [19] Arroyo-Valles, R. S. Maleki and G. Leus, A censoring strategy for decentralized estimation in energy-constrained adaptive diffusion networks. IEEE 14th Workshop Signal Process. Adv. Wireless Commun., 155-159, 2013.
- [20] R. Jiang, B. Chen, Fusion of censored decisions in wireless sensor networks. IEEE Trans. Wireless Commun., 4 (6), 2668-2673, 2005.
- [21] Z. Wang, Z. Yu, Q. Ling, D. Berberidis and G. B. Giannakis, Decentralized rls with data-adaptive censoring for regressions over large-scale networks. IEEE Trans. Signal Process., 66 (6), 1634-1648, 2018.
- [22] H. Zhu, H. Qian, X. Luo and Y. Yang, Adaptive queuing censoring for big data processing. IEEE Signal Process. Lett., 25 (5), 610-614, 2018.
- [23] A. O. Sarp, E. C. Mengüç, M. Peker and B. Ç. Güvenç, Data-adaptive censoring for short-term wind speed predictors based on MLP, RNN, and SVM. IEEE Systems Journal, 16 (3), 3625-3634, 2022.

- [24] J. Ferreira, M. Mendonca and P. S. Diniz, Data selection in neural networks. *IEEE Open Journal of Signal Processing*, 2, 522-534, 2021.
- [25] M. J. M. Spelta and W. A. Martins, Normalized lms algorithm and data-selective strategies for adaptive graph signal estimation. *Signal Process.*, 167, 107326, 2020.
- [26] K. L. Yin, Y. F. Pu and L. Lu, Censored regression distributed functional link adaptive filtering algorithm over nonlinear networks. *Signal Process.*, 190, 108318, 2022.
- [27] Y. Eren, B. Ç. Güvenç and E. C. Mengüç, Online censoring based acoustic feedback cancellation for wearable hearing aids. 30th Signal Processing and Communications Applications Conference (SIU), 1-4, Karabük, Türkiye, 2022.
- [28] E. C. Mengüç, S. Çınar, M. Xiang and D. P. Mandic, Online censoring based weighted-frequency fourier linear combiner for estimation of pathological hand tremors. *IEEE Signal Process. Lett.*, 28, 1460-1464, 2021.
- [29] P. S. R. Diniz, On data-selective adaptive filtering. *IEEE Transactions on Signal Processing*, 66 (16), 4239-4252, 2018.
- [30] A. Stott, S. Kanna and D. P. Mandic, Widely linear complex partial least squares for latent subspace regression. *Signal Processing*, 152, 350-362, 2018.
- [31] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity Widely Linear and Neural Models*. United Kingdom: Wiley, 2009.
- [32] T. Adalı, S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*. Wiley: IEEE Press, 2010.
- [33] Y. Xia, C. C. Took and D. P. Mandic, An augmented affine projection algorithm for the filtering of noncircular complex signals. *Signal Process.*, 90, 1788-1799, 2010.
- [34] B. Jelfs, D. P. Mandic and S. C. Douglas, An adaptive approach for the identification of improper complex signals. *Signal Process.*, 92, 335-344, 2012.
- [35] T. Adalı, P. J. Schreier and L. L. Scharf, Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Trans. Signal Process.*, 59 (11), 5101-5125, 2011.
- [36] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex- Valued Data: The Theory of Improper and Noncircular Signals (1. Basım)*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [37] E. C. Mengüç and N. Acır, An augmented complex-valued Lyapunov stability theory based adaptive filter algorithm. *Signal Processing.*, 137, 10-21, 2017.
- [38] E. C. Mengüç and N. Acır, An augmented complex-valued least-mean kurtosis algorithm for the filtering of noncircular signals. *IEEE Transactions on Signal Processing*, 66 (2), 438-448, 2018.
- [39] B. Widrow, J. McCool and M. Ball, The complex LMS algorithm. *Proceedings of the IEEE*, 63 (4), 719-720, 1975.
- [40] A. Tarighat and A. H. Sayed, Least mean-phase adaptive filters with application to communications systems. *IEEE Signal Processing Letters.*, 11 (2), 220-223, 2004.
- [41] S. Haykin, *Adaptive Filter Theory (4. Basım)*. Prentice-Hall, Upper Saddle River: Pearson Education, 2002.
- [42] A. Khalili, A. Rastegarnia, W. M. Bazzi and Z. Yang, Derivation and analysis of incremental augmented complex least mean square algorithm. *IET Signal Process.*, 9 (4), 312-319, 2014.
- [43] B. Picinbono and P. Chevalier, Widely linear estimation with complex data. *IEEE Trans. Signal Process.*, 43 (8), 2030-2033, 1995.
- [44] E. C. Mengüç and N. Acır, Complex-valued least mean kurtosis adaptive filter algorithm. 23rd Signal Process. Commun. Appl. Conf., 325-328, 2016.
- [45] A. Khalili and A. Rastegarnia, Tracking analysis of augmented complex least mean square algorithm. *Int. J. Adapt. Control Signal Process.*, 30 (1), 106-114, 2016.
- [46] A. Hakkarainen, J. Werner, K. R. Dandekar and M. Valkama, Analysis and augmented spatial processing for uplink OFDMA MU-MIMO receiver with transceiver I/Q imbalance and external interference. *IEEE Trans. Wireless Commun.*, 15 (5), 3422-3439, 2016.
- [47] Z. Shan and T. S. P. Yum, A conjugate augmented approach to direction of-arrival estimation. *IEEE Trans. Signal Process.*, 53 (11), 4104-4109, 2005.
- [48] P. Chevalier and A. Blin, Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences. *IEEE Transactions on Signal Processing*, 55 (11), 5323-5336, 2007.
- [49] S. Javidi, D. P. Mandic and A. Cichocki, Complex blind source extraction from noisy mixtures using second-order statistics. *IEEE Trans. Circuits Syst. I*, 57 (7), 1404-1416, 2010.
- [50] C. Park, C. C. Took and D. P. Mandic, Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks. *IEEE Trans. Neural Syst. Rehab. Eng.*, 22 (1), 1-10, 2014.
- [51] H. Li, N. M. Correa, P. A. Rodriguez, V. D. Calhoun and T. Adalı, Application of independent component analysis with adaptive density model to complex-valued fMRI data. *IEEE Transactions on Biomedical Engineering*, 58 (10), 2794-2803, 2011.
- [52] Y. Xia, S. C. Douglas and D. P. Mandic, Adaptive frequency estimation in smart grid applications: exploiting noncircularity and widely linear adaptive estimators. *IEEE Signal Process.*, 29 (5), 44-54, 2012.
- [53] Y. Xia and D. P. Mandic, Widely linear adaptive frequency estimation of unbalanced three-phase power systems," *IEEE Transactions on Instrumentation and Measurement.*, 61 (1), 74-83, 2012.
- [54] S. Javidi, S. L. Goh, M. Pedzisz and D. P. Mandic, The augmented complex least mean square algorithm with application to adaptive prediction problems. *Proc. 1st IARP Workshop Cogn. Inform. Process.*, 54-57, 2008.

- [55] S. Douglas, Widely-linear recursive least-squares algorithm for adaptive beamforming. Proc. the IEEE Int. Conf. Acoustics, Speech and Signal Process., 2041-2044, 2009.
- [56] B. Jelfs, D. P. Mandic and S. C. Douglas, An adaptive approach for the identification of improper complex signals. Signal Process., 92, 335-344, 2012.
- [57] C. Jahanchahi, S. Kanna and D. P. Mandic, Complex dual channel estimation: Cost effective widely linear adaptive filtering. Signal Process., 104, 33-42, 2014.
- [58] A. H. Sayed, Fundamentals of Adaptive Filtering (1. Basım). NJ: Wiley, 2003.
- [59] E. C. Mengüç, M. Xiang and D. P. Mandic, Online censoring based complex-valued adaptive filters. Signal Processing, 200, 108638, 2022.
- [60] Kaggle, Fifa 19 Dataset. <https://www.kaggle.com/karangadiya/fifa19>, Accessed 16 February 2022.
- [61] UC Irvine Machine Learning Repository, Block Feedback Dataset. <https://archive.ics.uci.edu/ml/datasets/BlockFeedback>, Accessed 8 May 2022.

