

Hadoop ile Kaos Temelli FCW Optimizasyon Algoritmasının Analizi

An Analysis of Chaos-Based the FCW Optimization Algorithm by Hadoop

Güngör YILDIRIM¹, Galip AYDIN¹, Hasan ALLI², Yetkin TATAR¹

¹Bilgisayar Mühendisliği Bölümü

Fırat Üniversitesi

gyildirim23@gmail.com, galipaydin@gmail.com, ytatar@gmail.com

²Makine Mühendisliği Bölümü

Fırat Üniversitesi

halli@firat.edu.tr

Özet

Optimizasyon ve dağıtık hesaplama günümüzün en enerjik araştırma konularındandır. Her iki alanda da her geçen gün yeni yöntemler ve teknikler kendini göstermeye devam etmektedir. Optimizasyonda “Kaos Tabanlı Optimizasyon Algoritmaları”, dağıtık ve paralel hesaplama teknolojilerinde ise “Hadoop Projesi” son zamanlarda kendilerinden çokça bahsettirmektedirler. Kaos optimizasyon algoritmalarının en temelinde “Birincil Taşıyıcı Dalga (FCW)” yaklaşımı bulunmaktadır. Literatürde bu teknik, sonuçsal bakımından çok başarılı bir yaklaşım olarak görülmemektedir. Bu çalışmada, Hadoop ve MapReduce teknolojilerinin güç ve esnekliği kullanılarak, FCW yaklaşımı tekrar analiz edilmiştir. Bunun için kaotik harita fonksiyonları yardımıyla oluşturulmuş büyük bir veri seti, Hadoop ve içinde FCW yaklaşımı barındıran MapReduce yazılım modeli ile test edilmiş ve sonuçlar paylaşılmıştır. Kullanılan karmaşık nonlineer test fonksiyonları sayesinde bu yaklaşımın söz konusu teknolojiler ile birlikte çok da başarısız olmadığı gösterilmiştir.

Abstract

Optimization and distributed computing fields are among the most energetic research subjects. Today many new methods and technics have been suggested in the two fields. In the optimization “Chaos based algorithms”, and in the distributed systems “Hadoop Project” have drawn attention in this context. “First Carrier Wave Algorithm - FCW” based on chaos is one of the basic approaches and it is known as a weak technic with respect to finding global minimum. In this study, the FCW technic was reexamined by Hadoop and MapReduce. For this a big data set created by well-known chaotic maps was used. Then the set was analyzed by Hadoop and a MapReduce model including the FCW codes and the results were shared. With complex nonlinear benchmark functions, the FCW technic was tested and its results to be considered successful were shown.

1. Giriş

Günümüzde bulgusal (heuristic) optimizasyon teknikleri önemli araştırma alanlarından biridir. Son yıllarda kaos temelli optimizasyon teknikleri bu alanda dikkatleri üzerine

çekmektedir. Doğası gereği lokal minimumlardan kaçabilme yeteneği, hızlı ve esnek olması gibi faktörlerden dolayı optimizasyon problemlerinin çözümünde incelenmeye değer bir konu olmaya devam etmektedir [1]. Bilindiği üzere optimizasyon problemlerinde ana amaç global minimumu (veya maksimumu) bulmaktır. Nonlineer fonksiyonlar yapılarında birçok minimum barındırabilirler. Bunların bir çoğu lokal minimumlardır ve önerilen yaklaşımlardan bu lokal minimumları aşip global minimumlara ulaşılması beklenir.

Kaos temelli optimizasyon algoritmalarının (KOA) işleyişlerinde, rastgele (random) fonksiyonlardan elde edilen değerler yerine kaotik harita fonksiyonlarından (Chaos Maps) elde edilen değerler kullanılır [2]. Kaotik haritaların ürettiği bu değerler, ilgili algoritmaya lokal noktalardan kaçabilme yeteneği kazandırabilmektedir. Ancak doğal olarak tek başına kaotik dizilerin kullanılması bu tekniği başarılı kılmamaktadır. Bu nedenle bir çok iyileştirme çalışmaları yapılmıştır ve yapılmaya devam etmektedir [3,4,5]. Özellikle bu alanda diğer teknikler ile beraber çalışma esasına dayalı hibrid teknikler ön plana çıkmaktadır.

Bilgisayar bilimleri konuları içinde üzerinde birçok araştırmanın yapıldığı bir diğer alan paralel ve dağıtık hesaplama teknolojileridir. Bu alan hem akademik hem de ticari manada etkinliğini sürdürmeye devam etmektedir. Son yıllarda bu alanda isimleri sıkça ön plana çıkan teknolojiler hiç şüphesiz “Hadoop” ve “MapReduce” dur. Hadoop, dağıtık hesaplama için geliştirilen, ölçeklenebilir, sağlam ve basit programlama modelleri ile kullanılabilen yazılımlar bütünüdür [6]. Yapısında bir çok alt modül barındırır. Açık kaynak kod olması birçok projede yer almasına da ön ayak olmuştur. Temel olarak büyük veri setlerinin, “commodity hardware” olarak bilinen bilgisayar kümeleri üzerinde işlenmesini sağlar. Yapısında kendisine en büyük avantajı sağlayan iki önemli özelliği barındırır. Bunlardan ilki kullandığı dosya sistemi HDFS, diğeri Google firması tarafından geliştirilmiş MapReduce yazılım modelidir. MapReduce modeli, programlama esnasında gerçekleştirilecek olan “Map” ve “Reduce” fonksiyonları üzerine çalışmasını inşa etmektedir. Map fonksiyonu büyük veri setinin daha ufak parçalara ayrılarak mevcut paralel sistem üzerine dağıtılmasını sağlar. Reduce fonksiyonu ise elden edilen ara sonuçların belirlenen stratejiye göre analizini yaparak uygun bir şekilde bu

sonuçların küçültülmesini ve sıralamasını kotarır. Çok özellikli ve maliyetli sunucular yerine sıradan kümesel makinelerin kullanıldığı bu teknolojiler sayesinde hem dinamik hem de düşük maliyetli paralel ve dağıtık sistem uygulamaları geliştirmek mümkün olmaktadır.

Çalışmanın bundan sonraki bölümleri sırası ile literatür özeti, kaos temelli optimizasyon algoritmaları ve birincil taşıyıcı dalga (FCW) tekniğinin izahatı, kaotik haritaların tanıtımları, Hadoop teknolojisinin özellikleri ve çalışma mantığı, MapReduce yazılım modelinin açıklanması, önerilen uygulamanın gerçekleştirilmesi ve sonuçlar şeklindedir.

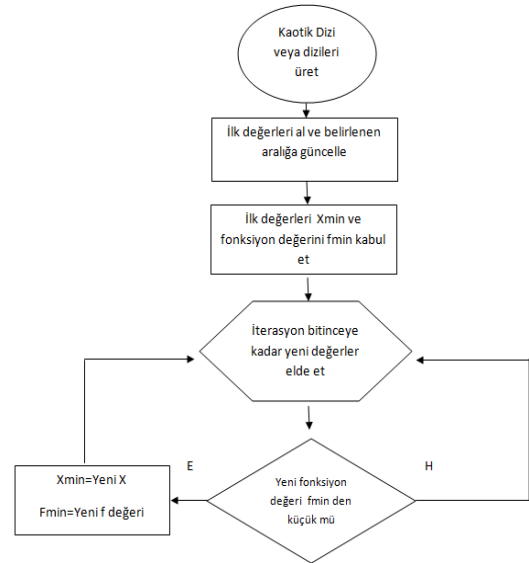
2. Literatür Özeti

Yuan ve arkadaşları kaos optimizasyon algoritmalarından iyileştirilmiş “Değişebilir Ölçekli Kaos Algoritması (MPCOA)” ’nı önermiş ve bunu harmony search ve simplex metodu gibi diğer tekniklerle birleştirerek sistemin daha geliştirilebilir olmasını sağlamaya çalışmışlardır. Bu çalışmada esasında çoklu FCW algoritması temeline dayalı olmakla birlikte bu geliştirilen tekniğin başlangıç şartlarına bağlı olmadığı ifade edilmiştir[1]. Bir diğer çalışmada farklı kaotik haritaların arama hızları ve olasılık yoğunluk fonksiyonuna (PDF) göre değerlendirmeleri gerçekleştirilmiştir. Bu çalışmada birçok test fonksiyonu için söz konusu kriterler değerlendirilmiştir. Temel kaos algoritmalarının incelendiği ve “Geliştirilmiş Periyodik Arama temelli Hibrid Kaotik Algoritması” adlı yeni bir yaklaşımın önerildiği çalışmada [2] dört farklı test fonksiyonuna göre sonuçlar analiz edilmiştir. Burada FCW algoritmasının nasıl geliştirilebileceği hakkında yapılan çalışmalar hakkında temel bilgiler sunulmuştur. Gandomi ve arkadaşları [3] son yıllarda adı duyulan meta sezgisel algoritmalarından ateşböceği algoritması (FA) ile kaotik yaklaşımları birleştirmiş ve sonuçlarını incelemiştir. Çalışma sonunda bazı kaotik haritalara dayalı FA 'ların standart FA 'ya göre daha iyi sonuç verdiklerinin gözlemlendiği belirtilmiştir. Tavazoei ve arkadaşları farklı kaotik haritalar kullanan arama sistemlerinin nonlineer optimizasyon problemlerinde verdikleri sonuçları analiz etmişlerdir [4]. Yarasa algoritması (BA) ile kaos yaklaşımını birleştirip, BA 'nın farklı varyasyonlarının farklı kaotik haritalarla davranışlarının incelendiği çalışmada ise [5] etkin sonuçların alındığı ifade edilmektedir. MapReduce modelinin esnekliğinin ve gelişiminin anlatıldığı makalede [7] mapreduce ile yapısının çalışma şekli açıklanmıştır. Ayrıca makalede Google firmasının on binin üstünde mapreduce modeli analiz yapıldığı ifade edilmiştir. Mapreduce modeli yaklaşımının on binlerce makine üzerinde terabaytlar boyutundaki verileri nasıl analiz edebildiğini açıklamış ve bununla alakalı bir gerçekleştirme yapmışlardır. Bu çalışmada ayrıca sistemde oluşabilecek hataların analizi ve tolerans yaklaşımları açıklanmıştır.

3. Temel Kaos Algoritmaları ve FCW

Kaos temelli optimizasyon algoritmalarında temel prensip, bilindik yaklaşımlarda sıkça tercih edilen rastgele (random) değişkenler yerine, kaotik harita olarak adlandırılan fonksiyonlardan elde edilen değerlerin kullanılmasıdır. Belirli bir iterasyon boyunca kullanılan kaotik değişkenlerin bütünü bir kaotik diziye temsil etmektedir. Kaotik dizilerin kullanılması

amaçları, doğasında bulunan esneklik sayesinde, global minimumu arama sürecinde, algoritmaya lokal minimumlardan kaçabilme yetisi kazandırmasıdır. Ancak pratik manada standart bilgisayar konfigürasyonlarında bu temel yaklaşımın çok başarılı olmadığı birçok çalışmada ifade edilmiştir[1,2]. Kaos temelli uygulamalarda prensipte dört yaklaşım vardır; Birincil Taşıyıcı Dalga yaklaşımı (FCW), İkincil Taşıyıcı Dalga Yaklaşımı (SCW), Gradyent Temelli Kaos Algoritmaları ve Kaotik-Metasezgisel Yaklaşımlar [2]. Genel çalışma mantığı Şekil 1 'de gösterilen FCW, kaos temelli yaklaşımların çekirdeğini oluşturmaktadır. Burada temel mantık kaotik haritalar yardımı ile elde edilen aday noktalar ile lokal minimumları geçip global minimuma ulaşmaktır. SCW ise FCW 'nin geliştirilmiş hali olup burada temel prensip, FCW 'den elde edilen minimum noktanın yeni bir aramanın başlangıç noktası olduğunu farz etmek ve bu noktanın baz alınarak ikinci bir kaotik dizi üretimi ile global minimumun aranmasıdır. Bu yaklaşımda bir “r” parametresi kullanılmaktadır. Bu değer aramanın yapılacağı yarıçapı göstermektedir. Ancak bu yaklaşım da FCW gibi çok başarılı sonuçlar vermemiştir. Gradyent temelli yaklaşımlar daha önce bahsedilen yaklaşımlara göre çok daha kararlı sonuçlar vermektedirler[1,2,4]. Bu yaklaşımda adında anlaşılacağı üzere gradyent temelli teknikler kaos temelli yaklaşımlarla hibrid olarak kullanılmışlardır. Bu yaklaşımlarda, genelde SCW yaklaşımlar tercih edilmiştir. Son yıllarda üzerinde sıkça araştırma yapılan farklı meta sezgisel yaklaşımların kaos temelli yapılarla beraber kullanıldığı bir çok çalışmayı da literatürde görmek mümkündür [3-5]



Şekil 1: Temel FCW 'nin akış diyagramı.

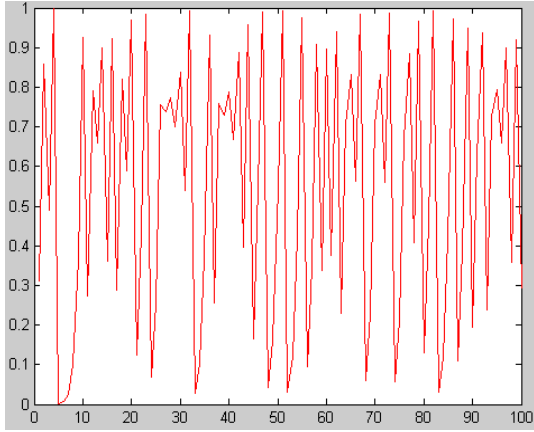
3.1. Kaotik Haritalar

Kaos temelli yaklaşımların temelinde kaotik haritalar olarak adlandırılan fonksiyonlar bulunmaktadır. Bu haritalar, kaotik temelli araştırmalarda ayrı bir inceleme alanı olarak dikkatleri üzerine çekmektedir. Birçok kaotik harita çalışması literatürde bulunmaktadır. Çok kullanılan bazı kaotik haritalar ve matematiksel fonksiyonları Çizelge 1 'de verilmiştir [2,4].

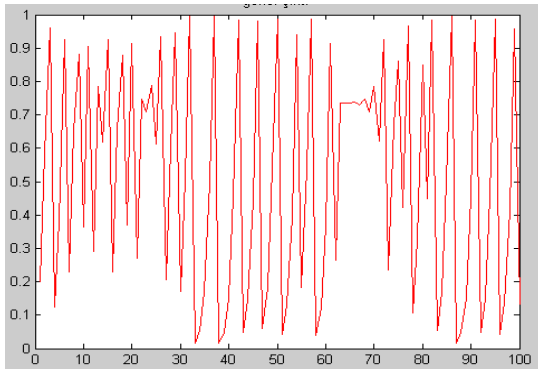
Çizelge 1: Sık kullanılan bazı kaotik haritalar

Adı	Fonksiyonu	Parametreler
Logistic Map [1]	$x_{n+1} = \mu x_n (1 - x_n)$	$0 < \mu \leq 4$
Sine Map [5]	$x_{n+1} = \frac{a}{4} \sin(\pi x_n)$	$0 < a \leq 4$
Chebyshev Map [3]	$x_{n+1} = \cos(k \cos^{-1} x_n)$	Sonuç [0 - 1] aralığında
Circle Map [5]	$x_{n+1} = x_n + 0.2 - \left(\frac{1}{4\pi}\right) \sin(2\pi x_n) \mod(1)$	
Singer Map [3]	$x_{n+1} = \mu(7.86x_n - 23.31x_n^2 + 28.75x_n^3 - 13.3x_n^4)$	$0.9 < \mu \leq 1.08$

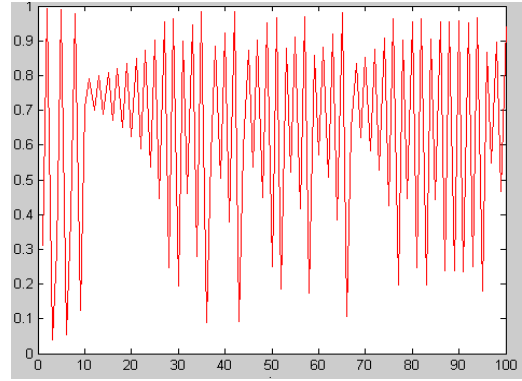
Bu haritalara ait grafikler Şekil 2-6 'da gösterilmektedir. Bu çalışmada literatürde sıkça kullanılan “Logistic Map” fonksiyonu, kaotik harita olarak tercih edilmiştir. Hazırlanan veri seti her ne kadar bu haritayı baz almış olsa da arzu edilen haritalama fonksiyonu ile farklı veri setleri rahatlıkla üretilebilmekte hatta çok boyutlu problemlerde amaç fonksiyona ait değişkenler için farklı kombinasyonlar denenebilmektedir.



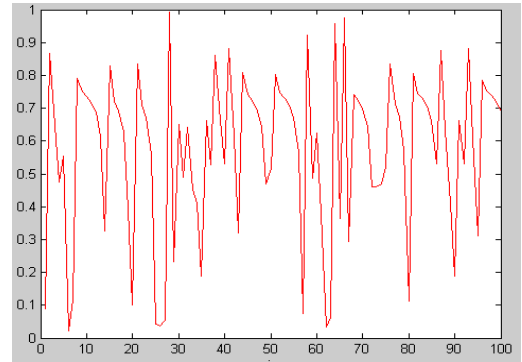
Şekil 2: Logistic Map



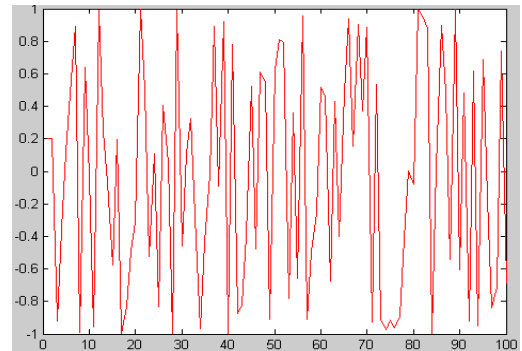
Şekil 3: Sine Map



Şekil 4: Singer Map



Şekil 5: Circle Map

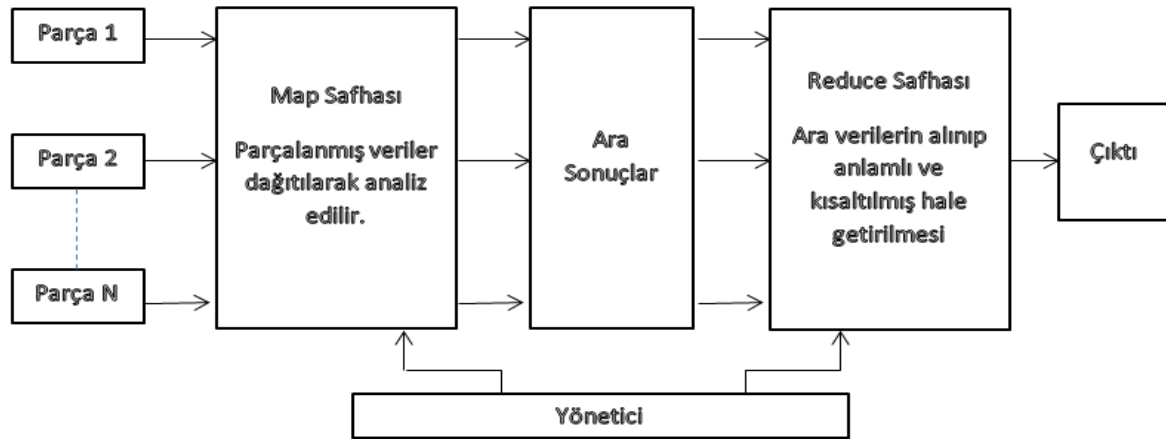


Şekil 6: Chebyshev Map

4. Hadoop ve MapReduce

Hadoop, bugün tanınmış birçok firma tarafından da tercih edilen ve desteklenen bir açık kaynak kod projesidir. Genel manada sıradan bilgisayarlardan oluşan kümesel yapılar üzerinde büyük boyutlara sahip veri setlerinin analiz edilmesini sağlar [6]. Bu yazılım “tekil (single node-mode)”, “sözde dağıtık (pseudo-distributed mode)” ve “tam dağıtık (fully-distributed mode)” modlarında çalışabilir. Hadoop projesi birkaç alt modül içermektedir. Bunlardan en önemlileri “Hadoop dağıtık Dosya Sistemi (HDFS)” ve “Hadoop MapReduce” dur. HDFS, dağıtık disklerin genel manada tek bir disk olarak çalışabilme yetisini sisteme kazandırır. Söz konusu büyük veri, çok daha küçük parçalara ayrılarak dağıtık sistem üzerinde paylaştırılır. Bunu yaparken

dağıtık sistemler için hayati kavramlarından bir olan “kopya saklama” özelliği de veri güvenliği açısından kullanılmaktadır. “NameNode” ve “DataNode” kavramları HDFS ‘nin temel yapı taşlarıdır [6-7]. NameNode HDFS ‘de bulunan tüm dizin yapısı bilgisini tutar ve yönetir. NameNode ‘da oluşabilecek bir hata istemin tümünü etkileyecektir. Bu nedenle ikincil veya yedek görev yapan NameNode ‘lar sitemde bulunmaktadır [6]. DataNode ‘lar HDFS içindeki verileri barındırmakla görevlidirler. Bir sistemde birden fazla DataNode bulunabilir. Bir istemcinin bir datanode ‘a ulaşması için öncelikle ilgili NameNode ‘dan gerekli bilgileri almış olması gerekmektedir.



Şekil 7: MapReduce genel çalışma prensibi

Çizelge 2: Kullanılan test fonksiyonları

Fonksiyon Adı	Fonksiyon Denklemi	Parametreler
Rastrigin Fonksiyonu	$f(x) = 10N + \sum_{i=1}^N [x_i^2 - 10\cos(2\pi x_i)]$	$-5.12 < x_i < 5.12$
Camel Fonksiyonu	$f(x) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_2^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$-200 < x_i < 200$
Sphere Fonksiyonu	$f(x) = \sum_{i=1}^N x_i^2$	$-100 < x_i < 100$
Nonlinear Test Fonk.	$f(x) = x_1^2 - 3x_1x_2 + 4x_2^2 + x_1 - x_2$	$-2 < x_i < 2$

Çizelge 3: Deney sonuçları

Fonksiyon	Gerçek Min. x 'ler	Gerçek fmin	Bulunan min. x 'ler	Bulunan fmin
Rastrigin Fonk. (N=2)	x1=0 x2=0	0	x1=0 x2=0	0
Rastrigin Fonk. (N=3)	x1=0 x2=0 x3=0	0	x1=-0.025 x2=-0.016 x3=-0.050	0.179
Camel Fonk.	x1=0.0898 x2=-0.712	-1.031	x1=0.080 x2=-0.720	-1.030
Sphere Fonk. (N=3)	x1=0 x2=0 x3=0	0	x1=-0.500 x2=-0.320 x3=-0.100	0.6
Nonlinear Fonk.	x1=-0.714 x2=-0.143	-0.28	x1=-0.714 x2=-0.143	-0.28

MapReduce daha öncede bahsedildiği gibi bir yazılım modelidir. Temel görevi dağıtık sistem üzerindeki çok büyük boyutlardaki verileri büyük kümesel yapılar üzerinde paralel olarak işlemektir. Ürettiği sonuçlar tekrar dağıtık sistem üzerinde tutulur. Bu yazılım modeli iki fonksiyondan meydana gelir; “Map” ve “Reduce”. Map fonksiyonun görevi, küçük parçalara ayrılmış verinin, kendi yapısındaki algoritmaya göre, dağıtık sistem üzerinde paralel olarak analiz edilmesini sağlamaktır. Reduce fonksiyonu ise map fonksiyonundan elde edilen ara verilerin küçültülerek daha anlamlı hale gelmelerini sağlar. Mapreduce ve HDFS aynı düğümler üzerinde çalışırlar. MapReduce, en temelde <anahtar,veri> modeli üzerine çalışmaktadır[7]. Genel yapı Şekil 7 ‘de gösterilmiştir.

5. Uygulama

Bu çalışmada amaç, kaos temelli FCW algoritmasının performansını MapReduce ile incelemek ve farklı bir teknolojiler ile optimizasyon tekniklerinin performansını test etmektir. Bunun için öncelikle “Logic Map” kullanan bir kod sayesinde 32 GB ‘lık bir veri seti oluşturulmuştur. Daha sonra FCW algoritmasını içeren MapReduce programı hazırlanmış ve Hadoop ‘un Pseudo-Distributed modunda test edilmiştir. Bu çalışmada daha genel bir izlenim elde edebilmek için iyi bilinen bazı karmaşık test fonksiyonlarından yararlanılmıştır. Bu fonksiyonlar ve parametre özellikleri Çizelge 2’de sunulmuştur. Bu çalışmaya ait deney sonuçları ise Çizelge 3’de verilmiştir.

6. Sonuçlar

Günümüzde gelişen bilgisayar teknolojileri ile birlikte optimizasyonda kullanılan bulgusal (heuristic) metotlar üzerindeki araştırmalarda hız kazanmıştır. Bu çalışmada, günümüzün etkileyici teknolojilerinden olan Hadoop ve MapReduce ile optimizasyonun ilgi çeken yöntemlerinden olan kaos optimizasyon algoritmasının çalışabilirlik analizi yapılmaya çalışılmıştır. Literatürde etkin sonuç vermeyen kaos temelli FCW algoritmasının Hadoop platformunda nasıl davranış göstereceğini incelemek için bir kaotik harita yardımıyla 32 GB ‘lık veri seti oluşturulmuş ve bu veri setinin işlenmesi için ise MapReduce programı yazılmıştır. Deney sonuçlarında elde edilen değerlerin başarılı sayılabilecek

veriler olduğu görülmüştür. Daha büyük veri setlerinde ve daha gelişmiş dağıtık sistemlerde bu değerlerin gerçek değerlere daha yaklaşıcağı anlaşılmaktadır.

7. Kaynaklar

- [1] Yuan, X., Yang Y., ve Wang, H., "Improved Parallel Chaos Optimization Algorithm", *Applied Mathematics and Computation*, 3590-3599, 2012, <http://dx.doi.org/10.1016/j.amc.2012.09.053>.
- [2] Velasquez, J.D., " An Enhanced Hybrid Chaotic Algorithm using Cyclic Coordinate Search and Gradient Techniques", *Revista de Ingeniera-Tecnica*, ISSN. 0121-4993, 45-53, 2010.
- [3] Gandomi, A.H., Yang X.S., Talatahari S., ve Alavi, A. H., " Firefly algorithm with chaos", *Commun Nonlinear Sci Numer Simulat*, 18, 89-98, 2013, <http://dx.doi.org/10.1016/j.cnsns.2012.06.009>.
- [4] Tavazoei M.S., Haeri M., " Comparison of different one-dimensional maps as chaotic search pattern in chaos optimization algorithms", *Applied Mathematics and Computation*, 187, 1076-1085, 2007, [doi:10.1016/j.amc.2006.09.087](http://dx.doi.org/10.1016/j.amc.2006.09.087).
- [5] Gandomi A.H., Yang X.S., " Chaotic bat algorithm", *Journal of Computational Science*, <http://dx.doi.org/10.1016/j.jocs.2013.10.002>
- [6] <http://hadoop.apache.org/>
- [7] Dean, J., Ghemawat, S., “MapReduce: Simplified Data Processing on Large Clusters” , *OSDI '04: 6th Symposium on Operating Systems Design and Implementation*, 137-149, 2004