SEYREK İŞARETLERİN GERİ KAZANIMINDA ORTALAMA HATA VE AYRIK FOURİER DÖNÜŞÜMÜ
AVERAGE ERROR IN RECOVERY OF SPARSE SIGNALS AND DISCRETE FOURIER TRANSFORM

Ayça Özçelikkale[1]*, Serdar Yüksel*[2]*, Haldun M. Özaktaş*[1]

[1]Elektrik Elektronik Mühendisliği Bölümü
Bilkent Üniversitesi, Ankara, Türkiye
ayca@ee.bilkent.edu.tr, haldun@ee.bilkent.edu.tr.
[2]Matematik ve İstatistik Bölümü
Queen's Üniversitesi, Kingston, Ontario, Kanada
yuksel@mast.queensu.ca

ÖZETÇE

Sıkıştırmalı algılama alanındaki yaklaşımlarda seyrek bir işaretin rasgele yerlerde alınan az sayıda ölçüm ile başarı ile geri kazanılabileceği gösterilmiştir. Ayrık Fourier dönüşümü (DFT) işaretin hangi bileşenlerinin sıfırdan farklı olduğundan bağımsız olarak en iyi performans garantilerini sağlayan dönüşümlerden biridir. Bu sonuç, performans ölçütü olarak yüksek olasılıkla işaret geri kazanımına dayanmaktadır. DFT'nin yüksek olasılık yerine ortalama hata ölçütü altında da genel bir eniyileyen dönüşüm olup olmadığı incelenmemiştir. Biz burada bu eniyileme sorusunu ele alıyoruz. Bu amaçla işareti rasgele bir süreç olarak modelliyoruz; ve işaretin değişinti matrisini işaretin seyrekliğinin bir ölçüsü olarak kullanan bir model öneriyoruz. Aksini ima eden çok sayıda sonuca rağmen DFT'nin her zaman en iyileyen dönüşüm olmadığını gösteriyoruz.

ABSTRACT

In compressive sensing framework it has been shown that a sparse signal can be successfully recovered from a few random measurements. The Discrete Fourier Transform (DFT) is one of the transforms that provide the best performance guarantees regardless of which components of the signal are nonzero. This result is based on the performance criterion of signal recovery with high probability. Whether the DFT is the optimum transform under average error criterion, instead of high probability criterion, has not been investigated. Here we consider this optimization problem. For this purpose, we model the signal as a random process, and propose a model where the covariance matrix of the signal is used as a measure of sparsity. We show that the DFT is, in general, not optimal despite numerous results that suggest otherwise.

1. GİRİŞ

Son zamanlarda işaret işleme alanında sıkıştırmalı algılama altında anılan yöntemler önem kazanmış; çeşitli uygulamalarda da başarı ile uygulanmıştır [1, 2]. Bu yöntemler seyrek işaretlere uygulanır. Seyrek işaretler, bir dönüşüm sonucunda (dalgacık ya da Fourier dönüşümü gibi) az sayıda sayı ile temsil edilebilen işaretlerdir. Araştırmacılar böyle bir işaretin uygun bir ölçüm dönüşümünden geçirildikten sonra rasgele yerlerde alınacak az sayıda ölçüm ile başarılı bir şekilde geri kazanılabileceğini göstermişlerdir. Burada özellikle etkileyici olan noktalardan biri, işaret ölçüm dönüşümünden geçirildikten sonra alınacak ölçümlerin yerlerinin dikkatle seçilmesinin gerekmemesidir. Bu işaret işleme alanındaki daha geleneksel yaklaşımlarımızdan farklıdır. Örneğin Fourier dönüşümünden geçtikten sonra sadece belli frekanslarda sıfırdan farklı bir değeri olan bir işareti az sayıda ölçüm ile geri kazanmak mümkündür ama ölçümlerin yeri özenle seçilmelidir.

Sıkıştırmalı algılama alanındaki sonuçlarda önemli bir kavram dönüşümlerle ilişkilendirilmiş olan uyumluluk parametresidir [1, 2, 3]. İki dönüşümün uyumluluğu —örneğin dikgen işaret dönüşümü ψ ile dikgen ölçüm dönüşümü ϕ'in uyumluluğu— $\mu = \max_{i,j} |u_{ij}|, U = \phi\psi$ ifadesi ile verilir. μ değeri küçük olduğu zaman uyumluluğun küçük olduğunu söyleriz. Uyumluluk $1/\sqrt{N} \leq \mu \leq 1$ aralığında değerler alabilir. Örneğin U eğer birim matris ise ($U = I$), $\mu = 1$ çıkar, ve uyumluluğun yüksek olduğu söylenir; eğer U ayrık Fourier dönüşümü (DFT) matrisi ise $\mu = 1/\sqrt{N}$ çıkar, ve uyumluluğun düşük olduğu söylenir. Görüldüğü üzere μ, U matrisinin sütunlarının ne kadar yayılmış olduğunun bir ölçüsüdür. μ küçüldükçe (uyumluluk azaldıkça), iyi performans garantileri elde edebilmek için daha az sayıda noktada ölçüm yapmak yeterli olmaya başlar.

DFT matrisinin (ya da dönüşümü temsil eden matri-

A. Özçelikkale'ye TÜBİTAK 2211 Yurt İçi Doktora Burs Programı ve TÜBİTAK 2214 Yurt Dışı Araştırma Burs Programı tarafından destek sağlanmıştır. A. Özçelikkale ve S.Yüksel kısmen Natural Sciences and Engineering Council of Canada tarafından desteklenmiştir. H. Özaktaş'a Türkiye Bilimler Akademisi tarafından kısmi destek sağlanmıştır.

978-1-4673-0056-8/12/$26.00 ©2012 IEEE

sin elemanlarinin hepsinin büyüklüğünün 1 olduğu başka bir dönüşümün) işaretin tam olarak hangi koordinatlarda seyrek olduğundan bağımsız olarak en iyi dönüşüm olarak karşımıza çıkması ilginç bir sonuçtur. Burada dikkat çekici olan unsurlardan biri performans kriterinin yüksek olasılıkla iyi işaret kestirimi sağlamak olmasıdır. Merak uyandıran bir soru DFT'nin performans kriteri olarak ortalama hata düşünüldüğünde eniyileyen dönüşüm olup olmadığıdır. Bu makalede bu soruya cevap vereceğiz.

Bu makalede olasılıksal bir işaret modeli kullanıyoruz ve en küçük ortalama karesel kestirici hatası (MMSE) üstünde yoğunlaşıyoruz. Önce işaretin değişinti matrisine dayanan bir seyreklik ölçütü öneriyoruz. Ardından sıkıştırmalı algılama sonuçlarının yarattığı beklentinin aksine genelde DFT'nin ortalama MMSE kriteri altında en iyi performansı veren dönüşüm olmadığını gösteriyoruz.

2. ÖLÇÜM MODELİ

Bu makalede aşağıdaki gürültülü ölçüm modeli düşünülmektedir:

$$y = Hx + n. \tag{1}$$

Burada $x \in \mathbb{C}^N$ bilinmeyen girdiyi, $n \in \mathbb{C}^M$ ölçüm gürültüsünü, $y \in \mathbb{C}^M$ ölçüm sonucunu, $H \in \mathbb{R}^{M\times N}$ rasgele ölçüm matrisini temsil eder. x ve n sıfır ortalamalı öz karmaşık Gaussian vektörlerdir. x ve n istatistiksel olarak bağımsızdır. H burada işaretin hangi noktalarda ölçüleceğinin rasgele olarak seçilmesini modellemektedir. Bir bileşen diğer bileşenlerin ölçülmesinden bağımsız olarak p olasılıkla ölçülür. Dolayısıyla H'in satırları birim matristen alınıp alınmayacağına p olasılıkla karar verilmiş olan satırlardan oluşur.

Ölçüm gürültüsünün değişinti matrisi $K_n = E[n_i n_i\dagger] = \sigma_n^2 I$, $\sigma_n^2 > 0$ ile, bilinmeyen işaretin değişinti matrisi $K_x = E[xx^\dagger]$ ile temsil edilir. Bu matrisin özdeğer ayrışımını $K_x = U\Lambda_x U^\dagger \succeq 0$ şeklinde ifade ediyoruz. Burada U matrisi $N \times N$'lik birimcil bir dönüşümdür, özdeğerler köşegen $\Lambda_x = \text{çapraz}(\lambda_1, \dots, \lambda_N)$, $\lambda_1 \geq \lambda_2, \dots, \geq \lambda_N$ matrisinde yer alır, çapraz$()$ ifadesi köşegeninde verilen değerler olan bir matrisi ifade etmektedir. $\dagger$ matrisin devriğinin karmaşık eşleniğini ifade eder. Burada sıfırdan farklı olan özdeğerlerin indeks kümesi olarak $B = \{i : \lambda_i > 0\}$'yi tanımlamak faydalı olacaktır. Bu kümeye bağlı olarak U_B'yi U'dan indeksi B kümesi içerisindeki sütunları alarak oluşturulmuş $N \times |B|$ büyüklüğündeki matris olarak tanımlıyoruz. Benzer şekilde $\Lambda_{x,B}$ de Λ_x'ten indeksi B kümesi içerisindeki satır ve sütunları almak sureti ile oluşturulmuş matris olarak tanımlanıyor. Dolayısıyla özdeğer ayrışımı $K_x = U\Lambda_x U^\dagger = U_B \Lambda_{x,B} U_B^\dagger$ şeklinde de yazılabilir.

Modelimizde hem işaretlerin hem de ölçüm yapılacak yerlerin rasgele süreçler olarak modellendiğini vurgulamak istiyoruz. Gerektiğinde istatistiksel ortalamaların hangi değişkenlere göre alındığını vurgulamak için aşağıdaki ifadeleri kullanacağız: $E_H[.]$ istatistiksel ortalamanın rasgele ölçüm matrisine göre, $E_S[.]$ ise işaretlere göre alındığını gösteriyor. Makalemizde "ortalama MMSE" ifadesini hem ölçüm matrisi hem de işaret matrisi üzerinden alınmış istatiksel ortalama sonucunda ortaya çıkan hata için kullanıyoruz. İşaret kestilirken H matrisinin bilindiğini varsayacağız. $N \times N$ büyüklüğündeki DFT matrisini F ile göstereceğiz. Bu matrisin elemanları şu ifade ile verilir: $f_{tk} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}tk}$, $0 \leq t, k \leq N-1, \sqrt{-1} = j$.

2.1. Seyreklik ölçütü olarak özdeğer dağılımı

Biz burada işaretin değişinti matrisinin özdeğerlerinin dağılımını işaretin seyreklik ölçüsü olarak öneriyoruz. Eğer özdeğer dağılımı dengeli değilse (bazı özdeğerler çok yüksekken diğerleri düşük ise) bu işaretin seyrek olduğu şeklinde yorumlanır. Eğer tüm değerler birbirlerine yakın iseler bu işaretin seyrek olmadığı şeklinde yorumlanır. Burada dikkat çeken özel bir durum özdeğerlerin bir kısmının tam olarak 0 olduğu durumdur. Bu durumda bu özdeğerlere karşılık gelen rasgele değişkenlerin bir olasılıkla sıfır oldukları gösterilebilir. Kalan özdeğerler eşit ise bunların sayısı işaretin serbestlik derecesinin ölçüsü olarak kullanılabilir. Bu sıkıştırmalı algılama alanındaki seyreklik ölçüsü ile birebir örtüşmektedir: belirlenimci bir x'in seyreklik ölçüsü gerekirse uygun bir dikgen dönüşümden geçtikten sonra işaretleri ifade edebilmek için gereken en az bileşen sayısıdır. Bu bakış açısı özdeğerlerin dağılımının MMSE kriteri altındaki yorumu ile de uyumludur: özdeğerler bu işaret ailesinden gelen bir işaretin verilen belli bir hata seviyesinden daha büyük olmayan bir hata seviyesi ile ifade edilebilmesi için en az kaç tane değişken kullanılması gerektiğini belirler. Özdeğerlerin hepsinin sıfır olmadığı durumlarda da şöyle bir tanım yapılabilir: İşaretin toplam gücü P olsun: $\sum_{i=1}^{N} \lambda_i = P$. $D(\delta)$, belli bir $\delta \in (0, 1]$ değeri için şu eşitsizliği sağlayan $\sum_{i=1}^{D} \lambda_i \geq \delta P$ en küçük sayı olarak tanımlanabilir. Dolayısıyla δ bire yakın olduğunda, $D(\delta)$ değişinti matrisinin etkin kertesi, ve işaret ailesinin etkin serbestlik derecesi olarak düşünülebilir.

Bu yorum bilişim kuramındaki entropi kavramı ile de tutarlıdır. Entropi kavramı rasgele bir kaynaktaki belirsizliği nicelendirmek üzere önerilmiştir. Bir işaret kaynağındaki belirsizlik arttıkça kaynağın entropisi de daha yüksek değerler alır. Yukarıda tanımlanan Gaussian kaynağın entropisi şu ifade ile verilir: $h(x) = \log(|\pi e \mathbf{K}_x|)$ bit [4]. Dolayısıyla işaret kaynağının entropisi özdeğer dağılımı tarafından belirlenir:

$$h(x) \propto \log(|\mathbf{K}_x|) = \sum_{i}^{N} \log(\lambda_i). \tag{2}$$

Özdeğerler dağılımı yayvanlaştıkça (seyreklik azaldıkça) entropi artar. Tersine bazı değerlerde yüksek diğerlerinde düşük değerler almaya başladıkça (seyreklik arttıkça) azalır.

3. ENİYİLEME PROBLEMİ

$\mathbb{U}^{\mathbb{N}}$, $N \times N$ büyüklüğündeki birimcil matrislerin, $\{U \in \mathbb{C}^N : U^\dagger U = I\}$, kümesi olsun. Biz şu enküçültme problemi ilgileniyoruz:

$$\inf_{U \in \mathbb{U}^N} E_{H,S}[||x - E[x|y]||^2], \tag{3}$$

Benzer bir problem kanal kapasitesini enbüyütmek amacı ile [5]'de düşünülmüştür. Bu çalışmada vericinin işaret kaynağına ait değişinti matrisinin özdeğerlerinin dağılımını da şekillendirmesine izin verilmektedir. Sıkıştırmalı algılama alanında MMSE üstünde yoğunlaşan çalışmalar bulunmakla beraber, bu çalışmalarda alıcının kanal bilgisine sahip olmadığı varsayılan senaryolar altında ortalama hata değil yüksek olasılıkla iyi hata kestirimleri sağlayan metodlar geliştirmek üzerinde çalışılmıştır [6, 7]. Biz burada alıcının kanal bilgisine sahip olduğunu varsayarak ortalama hata üstünde özdeğer dağılımın değiştilemediği bir senaryo üstünde yoğunlaşıyoruz.

Eniyilenmesi amaçlanan fonksiyondaki işaretlere göre alınan istatiksel beklenti operasyonu şu şekilde ifade edilebilir [8, Ch2]:

$$\begin{aligned}
E_S[||x - E[x|y]||^2] &= \mathrm{tr}(K_x - K_{xy}K_y^{-1}K_{xy}^\dagger) \\
&= \mathrm{tr}(K_x) - \mathrm{tr}(K_x H^\dagger (H K_x H^\dagger + K_n)^{-1} H K_x) \\
&= \mathrm{tr}(\Lambda_{x,B}) \\
&\quad - \mathrm{tr}(\Lambda_{x,B} U_B^\dagger H^\dagger (H U_B \Lambda_{x,B} U_B^\dagger H^\dagger + K_n)^{-1} H U_B \Lambda_{x,B}) \\
&= \mathrm{tr}\,((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H^\dagger H U_B)^{-1}) \qquad (4)
\end{aligned}$$

Burada tr matrisin izini yani özdeğerlerinin toplamını veren operatörü göstermektedir. Sadeleştirme için uygun boyutlardaki herhangi bir M matrisi için geçerli olan $\mathrm{tr}(U_B M U_B^\dagger) = \mathrm{tr}(M U_B^\dagger U_B) = \mathrm{tr}(M)$ şeklindeki eşitlik kulllanılmıştır. (4)'ün elde edilmesi için Sheerman-Morrison-Woodbury eşitliği kullanılmıştır [9].

Farklı ölçüm senaryolarını k, $1 \leq k \leq 2^N$ ile listeleyelim. k'inci ölçüm senaryosuna karşılık gelen ölçüm matrisi H_k ile; bu ölçüm senaryosunun olasılığı p_k ile gösterilsin. Dolayısıyla enküçültelecek fonksiyon şöyle ifade edilebilir:

$$E_{H,S}[||x - E[x|y]||^2] \qquad (5)$$

$$= E_H[\mathrm{tr}\,((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H^\dagger H U_B)^{-1})] \qquad (6)$$

$$= \sum_k p_k \,\mathrm{tr}\,((\Lambda_{x,B}^{-1} + \frac{1}{\sigma_n^2} U_B^\dagger H_k^\dagger H_k U_B)^{-1}) \qquad (7)$$

Bu enküçültme problemi U değişkenin sürekli bir fonksiyonudur. U'nun birimcil matris olma kısıtı $\mathbb{C}^{N\times N}$'in kapalı ve sınırlı bir alt kümesini tanımladığından enküçültme değeri ulaşılabilirdir.

Öte yandan bu bir dışbükey eniyileme problemi değildir, çünkü değişken uzayı $\mathbb{U}^{\mathbb{N}}$ dışbükey bir uzay değildir. Elimizde iki tane birimcil matris olsun: $U_1, U_2 \in \mathbb{U}^{\mathbb{N}}$. Bu durumda genel olarak $\theta U_1 + (1-\theta)U_2 \notin \mathbb{U}^{\mathbb{N}}$, $\theta \in [0,1]$. Örneğin $N = 1$, $U_1 = 1, U_2 = -1, \theta U_1 + (1-\theta)U_2 = 2\theta - 1 \notin \mathbb{U}^1, \quad \forall\, \theta \in [0,1]$.

3.1. DFT eniyileyen matris değildir

Problemi yukarıdaki şekilde ifade etmek suretiyle bu enküçültme probleminin çözümünün sağlaması gerekli olan eniyileme koşullarını bulmayı kolaylaştırmış olduk. Bu koşullar problemin bazı özel durumlardaki çözümleri ile beraber [10]'da elde edilmiştir. İncelenmis olan özel durumlarda da DFT bu en iyileme probleminin çözümü olarak karşımıza çıkmaktadır. Yine [10]'da DFT'nin MMSE kriteri altında da yüksek olasılıkla iyi performans garantileri sağladığını gösterilmiştir. Şimdi DFT'nin bu eniyileme probleminin ortalama hata altında genel bir çözümü olmadığını bir karşı örnek göstermek yolu ile kanıtlayacağız.

Bu amaçla DFT'den farklı bir birimcil matris ile elde edilen ortalama hatanın DFT ile elde edilen hatadan daha küçük olduğu bir örnek vereceğiz. Örneğimizde şu parametreleri kullanıyoruz: $N = 3$, $\Lambda_x = \text{çapraz}(1/6, 2/6, 3/6)$, $K_n = I$. Birimcil matris U'yu şu matris olarak seçiyoruz:

$$U_0 = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \qquad (8)$$

Bu durumda değişinti matrisi şu hali alır:

$$K_x = \begin{bmatrix} 1/3 & 0 & 1/6 \\ 0 & 1/3 & 0 \\ 1/6 & 0 & 1/3 \end{bmatrix} \qquad (9)$$

Hata, alınan ölçüm sayısına bağlı bir fonksiyon olarak $J(U) = \sum_{M=0}^{3} p^M (1-p)^{3-M} e_M(U)$ şeklinde ifade edilebilir. Burada e_M, M tane ölçüm yapılan durumlardaki hataların toplamını ifade etmektedir. $U = U_0$ ve $U = F$ durumlarında elde edilen hatalara baktığımızda şu değerleri buluyoruz: $e_0(U_0) = e_0(F) = 1$, $e_1(U_0) = e_1(F) = 65/24$, $e_3(U_0) = e_3(F) = 61/84$, ve $e_2(U_0) = 409/168$, $e_2(F) = 465/191$. Burada farklı olan değerler $M = 2$ durumu için bulunan değerlerdir: $e_2(U_0) < e_2(F)$, $e_2(U_0) = 409/168 \approx 2.434524$ ve $e_2(F) = 465/191 \approx 2.434555$. Böylece U_0 ile DFT'den daha iyi hata değerleri elde edildiğini göstermiş olduk. Buradaki tartışmamız DFT matrisinin sütunlarının sırasının değiştirildiği durumu da kapsamaktadır. Bu durumda özdeğerler ile DFT matrisinin sütunları arasındaki eşleşme değişir. Yukarıda verilen hata değerleri bu durumlarda da geçerlidir.

4. SONUÇLAR

Bu makalede işaret kaynaklarındaki seyrekliği modellemek için sıkıştırmalı algılama alanında gözden kaçırılmış bir model sunduk. Değişinti matrisinin özdeğerlerine dayanan bu model bize daha önce iyi incelenmemiş bazı ilişkileri inceleyebilmenin yolunu açtı. Bu model altında rasgele alınmış ölçümler altında ortalama MMSE performansı ile ilgilendik. DFT'nin, sıkıştırmalı algılama alanındaki sonuçların yarattığı beklentiye, yüksek olasılıkla iyi performans garantileri sağlıyor olmasına ve bazı özel durumlarda eniyileyen dönüşüm olmasına rağmen her zaman eniyileyen dönüşüm olmadığını gösterdik. Bu sonuç ilerde çalışılacak yeni bazı soruları da gündeme getirmiştir. Genel durumda, eniyileyen dönüşüm matrisi ile elde edilen hata değerinin DFT ile elde edilenden ne kadar düşük olabileceği ilerde incelemeyi düşündüğümüz ilginç bir noktadır.

5. KAYNAKÇA

[1] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, June 2007.

[2] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, pp. 2845 –2862, Nov. 2001.

[3] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1 – 24, 2008.

[4] I. E. Telatar, "Capacity of multi-antenna Gaussian chan-

nels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[5] A. Tulino, S. Verdu, G. Caire, and S. Shamai, "The Gaussian erasure channel," in *IEEE International Symposium on Information Theory, 2007*, pp. 1721 –1725, June 2007.

[6] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *IEEE Transactions on Information Theory,*, vol. 55, pp. 4701–4714, Oct. 2009.

[7] M. Protter, I. Yavneh, and M. Elad, "Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3471–3484, July 2010.

[8] B. D. O. Anderson and J. B. Moore, *Optimal filtering*. Prentice-Hall, Englewood Cliffs, N.J. :, 1979.

[9] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Review*, vol. 23, no. 1, pp. 53–60, 1981.

[10] A. Özçelikkale, S. Yüksel, and H. M. Ozaktas, "Unitary precoding and basis dependency of MMSE performance for Gaussian erasure channels," 2011. `arXiv:1111.2451v1 [cs.IT]: http://arxiv.org/abs/1111.2451`.