

Universidade Federal de Campina Grande

Departamento de Engenharia Elétrica

Anotações sobre Amostragem Compressiva

Edmar Candeia Gurjão
ecandeia@dee.ufcg.edu.br

Outubro de 2009

Capítulo 1

Introdução

Este texto contém uma compilação de idéias relacionadas à amostragem compressiva, tradução feita pelo autor a partir de *Compressed Sensing*. Buscou-se reunir os principais resultados encontrados na literatura.

O número de amostras necessárias para reconstruir sem erro um sinal é ditado pela sua largura de banda. Sabe-se que alguns sinais tem uma estrutura, e dessa forma podem ser comprimidos eficientemente sem muita perda perceptível [1]. Antes de obter essa representação compacta, o sinal inicialmente é adquirido (amostrado), calcula-se todos os coeficientes dessa representação, codifica-se os maiores coeficientes e descarta-se os demais. A questão que se levanta é se não seria possível, já na aquisição do sinal, levar em conta esse descarte de coeficientes e obter somente os mais representativos. Essa pergunta é respondida pela amostragem compressiva, *Compressed Sampling* ou *Compressed Sensing* em inglês. Nesse texto usaremos o termo em Amostragem Compressiva e não serão apresentadas provas dos resultados retirados da literatura, visto que as mesmas poderão ser encontradas nas referências.

Para um sinal $f \in \mathbb{C}^N$, define-se a transformada de Fourier discreta como $\mathcal{F}[f] = \hat{f} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ da seguinte forma

$$\hat{f}(\omega) = \sum_{t=0}^{N-1} f(t)e^{-2\pi i\omega t/N}, \quad \omega = 0, 1, \dots, N-1.$$

Se forem dados os valores dos coeficientes $\hat{f}(\omega)$ para todas as frequências $\omega \in \mathbb{Z}_N$, então pode-se reconstruir $f(t)$ exatamente pela transformada inversa.

Supondo que se conhece apenas um conjunto dos coeficientes $\hat{f}|_{\Omega}$ amostradas de um subconjunto $\Omega \subsetneq \mathbb{Z}_N$, em geral não temos informação para reconstruir f , pois como essa função tem N graus de liberdade só conhecemos $|\Omega| < N$ desses graus.

Supondo agora que f tem suporte em um subconjunto T (a priori desconhecido) de \mathbb{Z}_N , isto é, assumo que f possa ser escrita como uma superposição de impulsos usando os elemento de T , isto é

$$f(t) = \sum_{\tau \in T} f(\tau)\delta(t - \tau)$$

Caso N seja primo, o teorema a seguir assegura a reconstrução de f com $|T|$ pequeno [2].

Teorema 1 *Seja um sinal com comprimento N , um inteiro primo. Seja Ω um subconjunto de $\{0, \dots, N-1\}$, e f um vetor com suporte em T tal que*

$$|T| \leq \frac{1}{2} |\Omega|.$$

Então f pode ser reconstruído univocamente de Ω e $\hat{f}|_{\Omega}$. Conversamente, se Ω não é o conjunto de todas as N frequências, então existem vetores distintos f e g tal que $|supp(f)|, |supp(g)| \leq \frac{1}{2} |\Omega| + 1$ e tal que $\hat{f}|_{\Omega} = \hat{g}|_{\Omega}$.

Se T e Ω são selecionados aleatoriamente e uniformemente, então é esperado que o teorema acima seja válido com probabilidade muito próxima de um.

Inicialmente deve-se observar que a norma l_p de um vetor \mathbf{a} é definida como

$$\|\mathbf{a}\|_p = \left(\sum_j |a_j|^p \right)^{\frac{1}{p}}.$$

Definindo $0^0 = 0$ tem-se a norma l_0 , que consiste basicamente na contagem dos elementos do vetor que são diferentes de zero. Quando $p = 1$ tem-se a norma l_1 dada por

$$\|\mathbf{a}\|_1 = \sum_j |a_j|.$$

Observe que na minimização da norma l_1 não há dependência com o tamanho dos vetores.

Em princípio, pode-se reconstruir f exatamente pela resolução de um problema de otimização dado por

$$\min_{g \in C^N} \|g\|_{l_0}, \quad \hat{g}|_{\Omega} = \hat{f}|_{\Omega}$$

sendo $\|g\|_{l_0}$ o número de termos não zero, $|\{t, g(t) \neq 0\}|$. Esse é um problema de otimização combinatorial e intratável mesmo para sinais de tamanho pequenos.

Uma estratégia computacionalmente mais eficiente para recuperar f de Ω e $\hat{f}|_{\Omega}$ é resolver

$$(P_1) \quad \min_{g \in C^N} \|g\|_{l_1} = \sum_{t \in Z_n} |g(t)|, \quad \hat{g}|_{\Omega} = \hat{f}|_{\Omega}$$

Fazendo $|T| \leq \alpha \cdot |\Omega| / \log N$ ($\alpha > 0$ é uma constante), a resolução do problema (P_1) permite a recuperação exata de f . Formalmente escreve-se

Teorema 2 *Seja $f \in C^N$ um conjunto discreto com suporte em conjunto desconhecido T , e escolha Ω de tamanho $|\Omega| = N_{\omega}$ aleatoriamente e uniformemente. Para um fator de precisão M , se*

$$|T| \leq C_M \cdot (\log N)^{-1} \cdot |\Omega|$$

então com probabilidade de no mínimo $1 - O(N^{-M})$, o problema de minimização (P_1) é único e igual a f .

Outra linha de trabalho é a que busca a decomposição de um sinal f usando formas de onda de um dicionário D . Esses trabalhos são diferentes do apresentado acima pois aqui deseja-se estimar um sinal a partir de dados incompletos.

Um fato interessante é que pode-se usar a aleatoriedade como um mecanismo de amostragem, pois pode-se extrair informação sobre um objeto de interesse a partir de um conjunto de observações selecionadas aleatoriamente.

Implementação da Amostragem Compressiva

Seja um vetor $\mathbf{x} \in \mathbb{R}^N$ e um conjunto \mathbf{y} de suas medidas dadas por

$$\mathbf{y}_k = \langle \mathbf{x}, \psi_k \rangle, \quad k = 1, \dots, K, \quad \mathbf{y} = \Psi \mathbf{x}.$$

Assim, adquire-se informação sobre o sinal \mathbf{x} projetando o mesmo sobre K vetores $\psi_k \in \mathbb{R}^N$. O caso de interesse aqui é quando $K \ll N$, ou seja, tem-se menos medidas que valores desconhecidos do sinal, e isso leva a um sistema de equações indeterminado. Entretanto, garantido que sinal tem algumas características é possível obter uma solução para esse problema, uma dessas características é apresentada a seguir.

Um vetor \mathbf{x} é S -esparço se seu suporte $\{i : x_i \neq 0\}$ tem cardinalidade menor ou igual a S . Foi mostrado na seção anterior que sempre é possível recuperar exatamente o vetor \mathbf{x} resolvendo o seguinte programa convexo

$$(P_1) \quad \min_{\tilde{\mathbf{x}} \in \mathbb{R}^n} \|\tilde{\mathbf{x}}\|_{l_1} \quad \text{sujeito a } \Psi \tilde{\mathbf{x}} = \mathbf{y}.$$

Re-escrevendo o Teorema 2 tem-se

Teorema 3 *Se \mathbf{x} é S -esparço e dados K coeficientes de Fourier com frequências selecionadas uniformemente de maneira aleatória, se*

$$K \geq C.S. \log N$$

então a minimização l_1 reconstrói x com uma probabilidade muito alta.

Um fato importante dos resultados apresentados é que um sinal \mathbf{x} pode ser exatamente reconstruído a partir de um função convexa que não assume qualquer conhecimento sobre o número de coordenadas de x diferentes de zero, suas localizações ou amplitudes.

Para entender a implementação da amostragem compressiva são necessários alguns conceitos, apresentados a seguir.

Esparsidade e Incoerência

Seja um vetor $\mathbf{x} \in \mathbb{R}^N$ os coeficientes de um sinal $f \in \mathbb{R}^N$ numa base ortonormal Ψ

$$f(t) = \sum_{i=1}^N x_i \psi_i(t) \quad t = 1, \dots, N$$

donde pode-se escrever $x = \Psi f$ sendo Ψ uma matriz N por N com as formas de onda ψ_i como linhas, ou de forma equivalente $f = \Psi^* x$.

Diz-se que um sinal f é esparso no domínio Ψ se a sequência de coeficientes tem suporte em um pequeno conjunto, e compressível se a sequência é concentrada próximo a um pequeno conjunto. Suponha que dispomos de dados subamostrados sobre f , da seguinte forma

$$\mathbf{y} = \Phi f.$$

Pode-se coletar informação parcial sobre \mathbf{x} via $\mathbf{y} = \Phi' x$, sendo $\Phi' = \Phi \Psi^*$. Neste cenário, pode-se recuperar f encontrando a decomposição com norma l_1 mínima

$$\min \|\tilde{x}\|_{l_1} \text{ tal que } \Phi' \tilde{x} = y.$$

Seja x_S um conjunto S esparço, ou seja, a aproximação obtida mantendo as S maiores entradas e fazendo as demais iguais a zero. O erro de recuperação não é maior que $\|x - x_S\|_{l_2}$.

Considere um processo de medição linear geral que calcula $M < N$ produtos internos entre \mathbf{x} e uma coleção de vetores $\{\phi_j\}_{j=1}^M$, de tamanho N , como $y_j = \langle \mathbf{x}, \phi_j \rangle$. Arrume as medidas y_j em um vetor $M \times 1$ e os vetores ϕ_j^T como linhas de uma matriz Φ , então tem-se

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s}$$

sendo $\Theta = \Phi \Psi$ uma matriz $M \times N$.

O processo de medida não é adaptativo, pois Φ é fixo e não depende do sinal \mathbf{x} . Os problemas são

- obter uma matriz de medida estável Θ tal que a informação importante em qualquer vetor K -esparço não seja prejudicada pela redução de dimensão de $\mathbf{x} \in \mathbb{R}^N$ para $\mathbf{y} \in \mathbb{R}^M$.
- obter um algoritmo de reconstrução \mathbf{x} de somente $M \approx K$ medidas \mathbf{y} .

A obtenção de matriz de medida Φ deve permitir a reconstrução do sinal \mathbf{x} de comprimento N a partir das $M < N$ medidas, o que aparenta ser um problema mal condicionado. Entretanto, se \mathbf{x} K -esparço e os K localizações dos coeficientes não zero são conhecidas o problema pode ser resolvido se $M \geq K$. Uma condição necessária e suficiente para que a reconstrução possa ser feita é que para qualquer vetor \mathbf{v} compartilhando as mesmas K entradas não zero e para um $\epsilon > 0$

$$1 - \epsilon \leq \frac{\|\Theta \mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leq 1 + \epsilon$$

Isto é, a matriz Θ preserva os comprimentos destes vetores K -esparços. Como as localizações normalmente não são conhecidas, uma condição suficiente para uma solução estável é que Θ satisfaça a condição acima par vetores $3K$ -esparços. Esta condição é conhecida como Propriedade da Isometria Restrita (RIP - do inglês restricted isometry property).

Uma condição equivalente conhecida como incoerência, requer que as linhas $\{\phi_j\}$ de Φ não possam ser esparçadamente representadas como colunas de $\{\psi_i\}$ de Ψ (ou vice-versa). Formalmente, dado um par de bases ortonormais $(\Phi, \Psi) \in \mathbb{R}^n$, define-se a coerência entre elas como

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq k, j \leq n} |\langle \phi_k, \psi_j \rangle|.$$

Ou seja, a coerência mede a maior correlação entre dois elementos de Φ e Ψ , e mostra-se que $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$.

Teorema 4 *Fixe $f \in \mathbb{R}^N$ e suponha que a sequência de coeficientes x de f na base Ψ é S -eparça. Selecione aleatoriamente e uniformemente m medidas no domínio Φ . Se*

$$m \geq C\mu^2(\Phi, \Psi)S \log n$$

para alguma constante C positiva, então consegue-se recuperar o sinal com alta probabilidade.

As condições acima pode ser obtidas se Θ for selecionada como uma matriz aleatória. Se os elementos $\theta_{i,j}$ são variáveis aleatórias independentes e identicamente distribuídas (iid) com função densidade de probabilidade Gaussiana com média zero e variância $1/N$, então as medidas \mathbf{y} são M combinações aleatoriamente ponderadas dos elementos de \mathbf{x} . Essa matriz tem as seguintes propriedades:

- é incoerente com a base $\Psi = I$ de impulsos com alta probabilidade. Mostra-se que essa matriz tem a propriedade RIP com alta probabilidade se $M \geq cK \log(N/K)$ sendo c uma constante pequena. Então um sinal K -eparço de comprimento N pode ser recuperado a partir de $M \geq cK \log(N/K) \ll N$ medidas aleatórias Gaussianas.
- É universal no sentido que $\Theta = \Phi\Psi$ será iid Gaussiana com alta probabilidade independente da escolha da base ortonormal Ψ .

O algoritmo de reconstrução deve utilizar as M medidas no vetor \mathbf{y} , a matriz de medição aleatória Φ e a base Ψ para reconstruir o vetor de comprimento N \mathbf{x} . Para tanto será usada a minimização da norma l_1 ou seja, deve encontrar \hat{s} tal que

$$\hat{s} = \operatorname{argmin} \|s'\|_1 \text{ tal que } \Theta s' = \mathbf{y}$$

Algumas dúvidas e observações:

- Como implementar a minimização da norma l_1 no Scilab ?
- Inserir exemplos de reconstrução pela norma l_1 como feito por Candes.

Decoding by Linear Programming

Seja o vetor de entrada $f \in \mathbb{R}^N$ e o vetor de saída de um canal $y = \mathbf{A}f + e$, sendo \mathbf{A} uma matriz $m \times n$ (assumido que $m > n$), e e um vetor arbitrário de erros com tamanho n . A matriz \mathbf{A} pode ser vista como um código linear, porém aqui as entradas dessa matriz podem ser números reais, e não pertencentes a um corpo finito como é comum em Teoria da Codificação.

Como é sabido, um código só pode detectar e ou corrigir um determinado conjunto de erros, o que define a sua capacidade de detecção e correção respectivamente. Caso não haja erro, pode-se recuperar f de $\mathbf{A}f$, porém consideramos aqui a presença de um erro e , e demonstra-se que para

$$\|e\|_{l_0} = |\{i : e_i \neq 0\}| \leq \rho m$$

pode-se recuperar f . A questão é para que valores de ρ essa reconstrução é possível em algoritmos práticos?

Para reconstruir f é suficiente reconstruir o vetor e , pois temos $\mathbf{A}f + e$. Isso será feito pela construção de uma matriz \mathbf{F} que aniquila \mathbf{A} , ou seja $\mathbf{F}\mathbf{A} = \mathbf{0}$, e então aplicar \mathbf{F} a y , obtendo

$$\tilde{y} = \mathbf{F}(\mathbf{A}f + e) = \mathbf{F}e,$$

e o problema então se reduz a reconstruir um vetor e das observações $\mathbf{F}e$.

Matching Pursuit

Uma das técnicas de reconstrução de sinais esparsos deriva da idéia de encontrar uma base para representar o sinal. Essa busca denominada de Perseguição de Base (do inglês, *Basis Pursuit*) é descrita a seguir. Após a descrição do problema geral é apresentada uma técnica específica denominada *Matching Pursuit* e em seguida um algoritmo para implementá-la.

Em [3] é estudada uma forma de construir uma base \mathcal{B} , não necessariamente ortogonal, para obter uma aproximação eficiente de um sinal f . O conjunto de N vetores de $\mathcal{B} = \{\mathbf{x}_{p_i}\}_{0 \leq i < N}$ são selecionados de um dicionário $\mathcal{D} = \{\mathbf{x}_p\}_{0 \leq p < d}$. Partindo desses vetores decompõe-se f na base

$$f = \sum_{m=0}^{N-1} a[p_m] \mathbf{x}_{p_m}.$$

Caso seja feita uma restrição para bases ortogonais mostra-se que a melhor escolha pode ser otimizada pela minimização de

$$C(f, \mathcal{B}) = \sum_{m=0}^{N-1} \Phi \left(\frac{|a[p_m]|^2}{\|f\|^2} \right)$$

sendo $\Phi(u)$ uma função concava. No caso da ausência de ortogonalidade, uma perseguição de base para a “melhor” base que minimiza a equação para $\Phi(u) = u^{1/2}$ e tem-se

$$C(f, \mathcal{B}) = \frac{1}{\|f\|} \sum_{m=0}^{N-1} |a[p_m]|$$

que consiste em minimizar a norma l_1 , e o problema pode ser re-escrito como um de programação linear.

Apesar de ser uma problema de programação linear, uma perseguição de base é computacionalmente complexa, pois faz-se uma minimização sobre todos os dicionários de vetores. Uma solução mais eficiente foi dada por Mallet e Zhan [4] que reduz a complexidade computacional e foi denominada *matching pursuit* apresentada a seguir.

Seja $\mathcal{D} = \{g_\gamma\}_{g \in \Gamma}$ um dicionário de $d > N$ vetores, tendo uma norma uniforme. Este dicionário inclui N vetores linearmente independentes que definem a base para o espaço \mathbb{C}^n .

Inicia-se pela projeção de f em um vetor $g_{\gamma_0} \in \mathcal{D}$ e em seguida calcula-se o resíduo

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + Rf$$

como Rf é ortogonal a g_{γ_0} tem-se

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + \|Rf\|^2.$$

Para minimizar $\|Rf\|$ deve-se escolher $g_{\gamma_0} \in \mathcal{D}$ que maximize $|\langle f, g_{\gamma_0} \rangle|$. Em alguns casos é mais eficiente encontrar g_{γ_0} tal que

$$|\langle f, g_{\gamma_0} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle f, d_\gamma \rangle|$$

sendo $\alpha \in (0, 1]$ um fator de otimização. A perseguição itera repetindo o procedimento acima sob o resíduo. Desse forma, sendo $R^0 f = f$ e $R^k f$ o resíduo de ordem k , a próxima iteração escolhe $g_{\gamma_m} \in \mathcal{D}$ tal que

$$|\langle R^k f, g_{\gamma_k} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle R^k f, d_\gamma \rangle|$$

e projeta $R^k f$ em g_{γ_k} tal que

$$R^k f = \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} + R^{k+1} f$$

e somando os resíduos de 0 a $M - 1$ tem-se

$$f = \sum_{k=0}^{M-1} \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} + R^M f$$

e mostra-se que $\|R^k f\|$ converge exponencialmente para zero quando k tende para infinito.

Em [5] apresenta-se o *matching pursuit* ortogonal (*Orthogonal Matching Pursuit*) definindo um sinal \mathbf{s} de dimensão d e com no máximo m componentes não zero, ou seja, um sinal m esparsos.

Partindo de uma sequência de N vetores $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ em R^d , que não dependem de \mathbf{s} são obtidas as medidas

$$\langle \mathbf{s}, \mathbf{x}_1 \rangle, \langle \mathbf{s}, \mathbf{x}_2 \rangle, \dots, \langle \mathbf{s}, \mathbf{x}_N \rangle$$

e a partir dessas medidas deseja-se recuperar o sinal \mathbf{s} . Para isso monta-se a matriz Φ cujas linhas são os vetores \mathbf{x}_i , denominada de matriz de medidas com colunas $\varphi_1, \dots, \varphi_d$.

Desse forma, a media pode ser escrita como

$$\mathbf{v} = \Phi \mathbf{s}$$

e o vetor \mathbf{v} será uma combinação linear das linhas de Φ . Na linguagem de aproximação esparsa \mathbf{v} tem uma representação de m termos sobre o dicionário Φ .

Para recuperar \mathbf{s} é necessário determinar que colunas de Φ participam da composição de \mathbf{v} . Para isso seleciona-se colunas de Φ de uma maneira gulosa, ou seja, em cada iteração escolhe-se as colunas de Φ que são mais fortemente correlacionadas com a parte restante de \mathbf{v} , então subtrai-se essa contribuição de \mathbf{v} e repete-se a iteração sobre o resíduo. Mostra-se que m iterações (esparsidade de \mathbf{s}) o algoritmo converge para uma boa aproximação.

O algoritmo 1 é uma repetição do apresentado em [5]. As estimativas $\hat{\mathbf{s}}$ fornecida por esse algoritmo para o sinal ideal tem os índices não zero listado em Λ_m . Os valores da estimativa $\hat{\mathbf{s}}$ são os componentes λ_j de \mathbf{x}_t .

Entradas:

- Uma matriz , de medição Φ , $N \times d$
- Um vetor N -dimensional \mathbf{v} de dados
- O nível m de esparsidade do sinal ideal

Saídas:

- Uma estimativa $\hat{\mathbf{s}}$ em R^d do sinal ideal
- Um conjunto Λ_m contendo m elementos de $\{1, \dots, d\}$
- Uma aproximação N dimensional \mathbf{a}_m do vetor \mathbf{v}
- Um resíduo N dimensional $\mathbf{r}_m = \mathbf{v} - \mathbf{a}_m$

Inicialização:

- Faça $\mathbf{r}_0 = \mathbf{v}$, $\Lambda_0 = \emptyset$, e o contador de inicialização $t = 1$

Iteração: Enquanto $t < m$ faça

1. Encontre o índice λ_t que resolve o problema de otimização

$$\lambda_t = \arg \min_{\mathbf{x}} \|\mathbf{v} - \mathbf{\Omega}_t \mathbf{x}\|_2$$

2. Amplie o conjunto índice e a matriz com os átomos escolhidos: $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$ e $\mathbf{\Omega}_t = [\mathbf{\Omega}_{t-1} \varphi_{\lambda_t}]$. Inicie com $\mathbf{\Omega}_0$ como a matriz vazia.
3. Calcule a nova aproximação dos dados e o novo resíduo

$$\mathbf{a}_t = \mathbf{\Phi}_t \mathbf{x}_t \text{ e } \mathbf{r}_t = \mathbf{v} - \mathbf{a}_t$$

4. Incremente t .

Algoritmo 1: Algoritmo OMP apresentado em [5].

Resumo de [?]

O método proposto codifica o sinal no domínio de tempo, ao invés dos parâmetros do modelo senoidal como os métodos atuais propõem.

Modelo Senoidal

O modelo senoidal foi inicialmente aplicado à análise e síntese de voz. Um sinal hamrônico $s(t)$ é representado com uma soma de um pequeno número de K senóides com amplitudes e frequências variantes no tempo, ou seja

$$s(t) = \sum_{k=1}^K \alpha_k(t) \cos(\beta_k(t))$$

sendo $\alpha_k(t)$ e $\beta_k(t)$ as amplitudes e fases instantâneas. Para estimar o modelo o sinal é dividido em quadros, e para cada um deles é calcula-se a representação em frequência. Dessa representação se obtém os picos espectrais e para cada pico no l -ésimo quadro determina-se a trinca $\{\alpha_{l,k}, f_{l,k}, \theta_{l,k}\}$ (amplitude, fase e frequência) para as K senóides.

Uma representação mais precisa do sinal de áudio pode ser obtida se for incluído um sinal de erro, porém isso não é feito nesse trabalho.

Como um sinal de áudio modelado por senóides é claramente esparso no domínio da frequência, a motivação é codificar esse sinal usando somente uma pequena parte das suas amostras.

Seja \mathbf{x}_l

Referências Bibliográficas

- [1] Emmanuel J. Candès. Compressed sampling. *Proceedings of the International Congress of Mathematicians*, 2006. Madrid, Spain.
- [2] Justin Romberg Emmanuel Candes and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Technical Report*, June 2005.
- [3] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [4] S. Mallat and A. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.
- [5] Joel A. Tropp and Anna C. Gilbert. Signal recovering from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, December 2007.