Anais do Congresso de Matemática Aplicada e Computacional
CMAC Sudeste 2013
ISSN 2237-7166

Escolha do Parâmetro de Relaxação e Implementações Paralelas do Método de Kaczmarz

Leonardo Bravo Estácio **Elias Salomão Helou**,
Instituto de Ciências Matemáticas e de Computação – ICMC – USP
Avenida Trabalhador São-Carlense, 400 – Centro – CEP: 13566-590 – São Carlos – SP
leobravo@icmc.usp.br elias@icmc.usp.br

15 de junho de 2013

Resumo: *O método de Kaczmarz é um algoritmo que soluciona sistemas lineares do tipo $Ax = b$ iterativamente através de projeções sobre hiperplanos. Strohmer e Vershynin [19] propuseram uma versão aleatória deste método e provaram possuir taxa de convergência esperada exponencial. Posteriormente, Eldar e Needell [4] sugeriram uma versão modificada do algoritmo de Strohmer e Vershynin, no qual a cada iteração é selecionada a projeção ótima à partir de um conjunto aleatório, utilizando para isto o lema de Johnson-Lindenstrauss. No entanto, em ambos os estudos, a escolha do parâmetro de relaxação não foi apresentada, o que pode melhorar muito a convergência dos algoritmos. Neste artigo apresentamos um estudo do parâmetro de relaxação para melhorar a convergência de ambos os algoritmos, e uma implementação em paralelo. Partimos dos resultados obtidos por Strohmer e Vershynin, porém desta vez considerando o parâmetro de relaxação e obtemos uma maneira de escolhê-lo. Além disso, através da paralelização do método, obtivemos grandes ganhos em sua velocidade de convergência.*

Palavras-chave: *Método de Kaczmarz, Algoritmos Iterativos, Escolha de Parâmetros, Métodos Numéricos, Algoritmos Paralelos*

1 Introdução

A solução de sistemas lineares da forma $Ax = b$, com $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ e $b \in \mathbb{R}^m$ aparecem com frequência em métodos numéricos, seja de forma direta ou como ferramenta intermediária na obtenção da solução desejada. Apesar da aparente simplicidade do problema de encontrar x satisfazendo a igualdade $Ax = b$, diversos fatores contribuem para que o uso de métodos diretos tradicionais seja inviável em aplicações que envolvam a resolução de grandes sistemas de equações lineares esparsos e de grande uso de memória. Para estas aplicações, são utilizados métodos iterativos [9, 7, 1, 16].

O método de Kaczmarz [13], também conhecido na literatura como Técnica de Reconstrução Algébrica (ART - *Algebraic Reconstruction Technique*), é um algoritmo iterativo para solução de sistemas lineares $Ax = b$. Devido à sua simplicidade e velocidade, vem sendo usado em uma variedade de aplicações partindo da tomografia e indo até o processamento digital de sinais [5, 14, 17]. O algoritmo consiste em uma série de projeções alternadas realizadas nos hiperplanos das equações do sistema.

Sejam $a_1, ..., a_m \in \mathbb{R}^n$ as linhas de A. O método de Kaczmarz clássico, de maneira cíclica, projeta ortogonalmente a estimativa corrente no hiperplano das soluções $\langle a_i, x\rangle = b_i$. Considerando cada iteração como uma única projeção ortogonal, o método pode ser descrito por

$$x_{k+1} = x_k + \lambda \frac{b_i - \langle a_i, x_k\rangle}{\|a_i\|_2^2} a_i^t, \tag{1}$$

 ISSN 2237-7166

onde x_k é a k-ésima iteração, $i = (k \bmod m) + 1$, $\|\cdot\|_2$ denota a norma Euclidiana ou norma espectral para vetores ou matrizes e λ o parâmetro de relaxação.

Embora esta técnica tenha sido usada na prática por bastante tempo, estimativas quantitativas de sua taxa de convergência foram difíceis de serem obtidas [3, 8, 11]. Pelo algoritmo, podemos perceber que a taxa de convergência depende da ordenação das linhas da matriz A e, portanto, linhas mal ordenadas podem conduzir a uma lenta convergência. Para superar esta dificuldade, as linhas da matriz A podem ser selecionadas de uma maneira aleatória. Foi observado que esta versão aleatória do algoritmo pode melhorar a taxa de convergência [12, 14], entretanto só recentemente resultados teóricos foram obtidos [15, 18, 19].

Uma variação do método clássico apresentada por Strohmer e Vershynin em [19] consiste em selecionar os hiperplanos para serem realizadas as projeções através de uma probabilidade proporcional a sua norma euclidiana. Sua abordagem pode ser descrita por

$$x_{k+1} = x_k + \lambda \frac{b_{p(i)} - \langle a_{p(i)}, x_k \rangle}{\|a_{p(i)}\|_2^2} a_{p(i)}^t,$$

onde $p(i)$ assume valores entre $\{1, ..., m\}$ com probabilidade $\frac{\|a_{p(i)}\|_2^2}{\|A\|_F^2}$ e $\|A\|_F$ denota a norma de Frobenius. Esta abordagem foi provada possuir limite exponencial na taxa de convergência de

$$\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2,$$

onde $R = \|A^{-1}\|_2^2 \|A\|_F^2$, x_0 é uma estimativa inicial arbitrária e $\mathbb{E}$ denota a esperança sobre a escolha das linhas. Isto assumindo que a matriz A é de posto completo de modo que $\|A^{-1}\| \stackrel{def}{=} \inf \{M : M\|Ax\|_2 \geq \|x\|_2, \forall\, x\}$ esteja bem definido.

Utilizando-se de tais resultados, foi possível acelerar a convergência do algoritmo através de uma estratégia de redução de dimensionalidade baseada no Lema de Johnson–Lindenstrauss [4]. Esta adaptação possui um caráter essencialmente paralelo na qual diversas projeções são calculadas (em um espaço de dimensão reduzida), mas somente a que mais avança na direção da solução é utilizada.

Vários trabalhos [2, 19] comentam que a taxa de convergência do método de Kaczmarz pode ser acelerada através da seleção de um bom parâmetro de relaxação. Entretanto, nenhum apresenta um critério para a sua seleção. A escolha de um parâmetro inapropriado pode reduzir a velocidade de convergência do algoritmo enquanto que bons valores para estes parâmetros resultam em um algoritmo altamente competitivo, em alguns casos sendo comparado favoravelmente com métodos bem estabelecidos como o dos gradientes conjugados [19].

2 Escolha do Parâmetro de Relaxação

Lema 2.1. *Seja* $Z_k := A_{i(k)}/\|A_{i(k)}\|_2$, *então o algoritmo* (1) *pode ser escrito como:*

$$x^{k+1} = x^k - \lambda \langle Z_k, x^k - x^* \rangle.$$

Utilizando o Lema 2.1, podemos chegar ao seguinte resultado:

$$\|x^{k+1} - x^*\|_2^2 = \|x^k - x^*\|_2^2 + \left(\lambda^2 - 2\lambda\right) \langle Z_k, x^k - x^* \rangle^2. \tag{2}$$

Considerando um passo anterior da Equação (2), chegamos a:

$$\|x^{k+1} - x^*\|_2^2 = \|x^{k-1} - x^*\|_2^2 + (\lambda^2 - 2\lambda)[\langle Z_{k-1}, x^{k-1} - x^* \rangle^2 + \langle Z_k, x^k - x^* \rangle^2]. \tag{3}$$

Na Equação (3), como x^k também depende de λ, escrevemos:

$$\langle Z_k, x^k - x^* \rangle = -\lambda \langle Z_{k-1}, x^{k-1} - x^* \rangle \langle Z_k, Z_{k-1} \rangle + \langle Z_k, x^{k-1} - x^* \rangle\ . \tag{4}$$

 ISSN 2237-7166

Substituindo a Equação (4) em (3), temos:

$$\|x^{k+1}-x^*\|_2^2 = \|x^{k-1}-x^*\| + (\lambda^2-2\lambda)\Big[\langle Z_{k-1}, x^{(k-1)}-x^*\rangle^2 + (-\lambda\langle Z_{k-1}, x^{k-1}-x^*\rangle\langle Z_k, Z_{k-1}\rangle + \langle Z_k, x^{k-1}-x^*\rangle)^2\Big]. \tag{5}$$

Portanto, a fim de minimizar $\|x^{k+1}-x^*\|_2^2$ teríamos que encontrar o minimizador de um polinômio de grau 4 em λ. Infelizmente, esta abordagem possui um horizonte muito curto e os valores obtidos para λ tendem a estar muito próximos de 1 e não possuem muita utilidade. Aumentar o número de iterações consideradas não é praticável, pois torna os cálculos muito custosos.

Podemos reescrever a Equação (2) da seguinte maneira:

$$\|x^{k+1}-x^*\|_2^2 = \left(1+\left(\lambda^2-2\lambda\right)\left\langle Z_k, \frac{x^k-x^*}{\|x^k-x^*\|_2}\right\rangle^2\right)\|x^k-x^*\|_2^2. \tag{6}$$

De (6), podemos facilmente obter, para $\lambda \in (0,2)$, a taxa de convergência

$$\mathbb{E}\|x^{k+1}-x^*\|_2^2 \leq \left(1+(\lambda^2-2\lambda)\hat{\kappa}(A)^{-2}\right)\mathbb{E}\|x^{(k)}-x^*\|_2^2, \tag{7}$$

onde $\hat{\kappa}(A) = \|A\|_F\|A^{-1}\|_2$ é o número de condição escalado da matriz A.

Este é o resultado obtido em [19], porém com a relaxação sendo levada em consideração. Obviamente, não se pode ver nenhuma utilidade para a relaxação aqui uma vez que o ótimo seria atingido com $\lambda = 1$. Porém, utilizando uma generalização destes resultados, obtemos o seguinte Lema:

Lema 2.2. *Suponha que Z_k segue a lei de probabilidade $P\left(Z_k = \frac{a_i}{\|a_i\|_2}\right) = \frac{\|a_i\|_2^2}{\|A\|_F^2}$, então temos:*

$$\mathbb{E}\langle Z_k, x\rangle\langle Z_k, y\rangle = \frac{1}{\|A\|_F^2}\langle Ax, Ay\rangle.$$

O que fizemos foi utilizar o Lema 2.2 em (5) e avaliar primeiro a esperança condicional a $Z_0, ..., Z_{k-1}$ e então condicional a $Z_0, ..., Z_{k-2}$, estimar as quantidades comparadas a $\|x^{k-1}-x^*\|_2^2$ e depois utilizar os resultados para estimar $\|x^{x+1}-x^*\|_2^2$ comparando a $\|x^{k-1}-x^*\|_2^2$. Obtendo:

$$\mathbb{E}^{k-1}\mathbb{E}^k\|x^{k+1}-x^*\|_2^2 \geq \left[1+\frac{\|A\|_2^2(\lambda^2-2\lambda)}{\|A\|_F^4}\left(2\|A\|_F^2+\lambda^2\rho-2\lambda\underline{\sigma}^2\right)\right]\|x^{k-1}-x^*\|_2^2 \tag{8}$$

onde $\rho := \max_{i=1...m}\frac{\|Aa_i\|_2^2}{\|a_i\|_2^2}$ e $\underline{\sigma}$ é o menor valor singular de A.

Ou seja, para escolhermos o parâmetro de relaxação λ, precisamos minimizar:

$$\left(\lambda^2-2\lambda\right)\left(2\|A\|_F^2+\lambda^2\rho-2\lambda\underline{\sigma}^2\right). \tag{9}$$

3 Simulação Numérica da Estimativa

Como um modelo eficiente de dimensão finita apropriado para o tratamento do problema de amostragem não uniforme, consideramos os polinômios trigonométricos discutidos por Feichtinger e Gröchenig em [10, 6] e utilizado por Strohmer [19] em suas simulações numéricas. O modelo consiste em solucionar o sistema linear:

$$Ax = b, \qquad \text{onde } A_{j,k} = \sqrt{w_j}e^{2\pi i k t_j}, \qquad b_j = \sqrt{w_j}f\left(t_j\right), \tag{10}$$

onde $\{t_j\}_{j=1}^m$ são valores uniformemente distribuídos, $w_j = \frac{t_{j+1}-t_{j-1}}{2}$, com $j = 1, \ldots, m$. $x = \{x_k\}_{k=-r}^r \in \mathbb{C}^{2r+1}$ é um vetor de valores complexos arbitrários e $f\left(t_j\right) = \sum_{k=-r}^r x_k e^{2\pi i k t_j}$.

ISSN 2237-7166

Utilizando o método de Kaczmarz cíclico e o aleatório de Strohmer no sistema linear (10), fixamos o número de iterações $k = 1400$ e analisamos o erro $\|x^k - x^*\|_2$ para diferentes valores do parâmetro de relaxação $\lambda \in (0, 2)$, obtendo a Figura 1. Onde a matriz $A \in \mathbb{C}^{m\times n}$ possui dimensões $m = 700$, $n = 101$, construída a partir de $r = 50$ e valores de t_j gerados através de uma função geradora de números pseudo-aleatórios uniformemente distribuídos dentro do intervalo $[0, 1]$ e ordenados de forma ascendente.

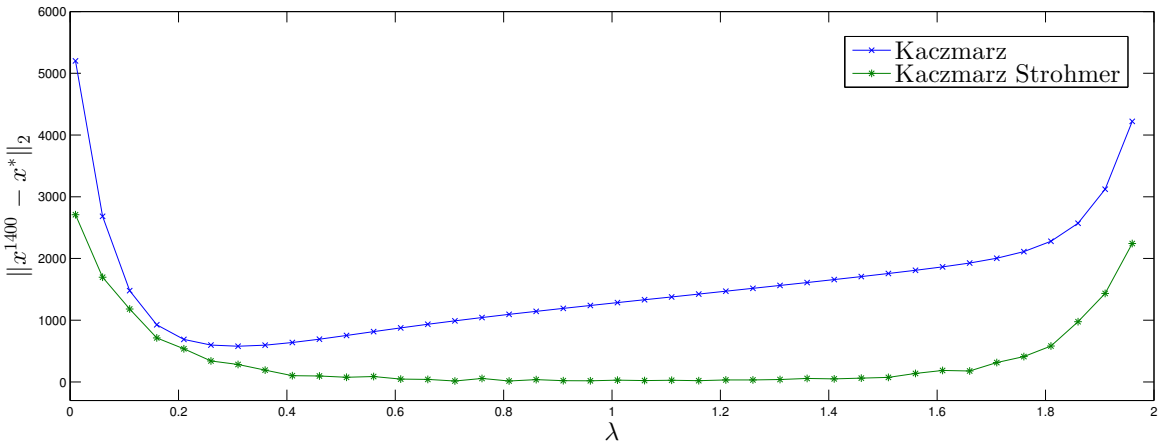


Figura 1: Comparação da taxa de convergência do algoritmo Kaczmarz cíclico e Kaczmarz aleatório de Strohmer em função da variação do parâmetro de relaxação. Os algoritmos foram aplicados ao problema de amostragem não uniforme descrita no texto.

Analisando a Figura 1, podemos ver que o valor do λ ótimo está entre $(0.6, 1.4)$ e qualquer valor dentro deste intervalo é uma boa aproximação.

Utilizando a fórmula (11), colocamos em escala $[0, 1]$ os valores de $\|x^k - x^*\|_2$ apresentados na Figura 1 e também os de nossa estimativa apresentada em (9) dado $\lambda \in (0, 2)$. Assim, construímos a Figura 2 avaliando estas escalas em função do parâmetro de relaxação.

$$s(x) = \frac{f(x) - \inf_{(0,2)} f(x)}{\sup_{(0,2)} f(x) - \inf_{(0,2)} f(x)} \tag{11}$$

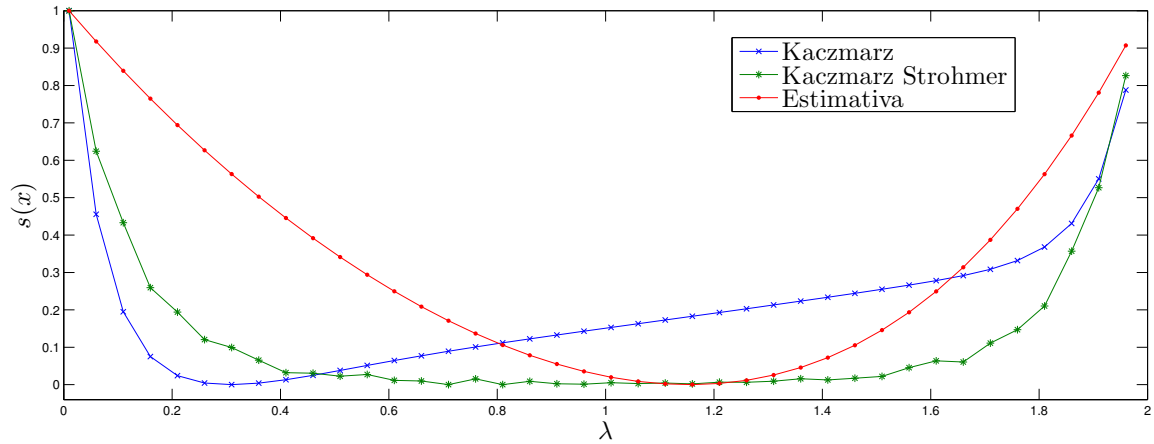


Figura 2: Valores de $\|x^k - x^*\|$ apresentando na Figura 1 e Valores de (9) com $\lambda \in (0, 2)$. Ambos em escala $[0, 1]$.

Desta maneira, encontrando o λ que minimiza a estimativa (9) na Figura 2, encontramos uma boa aproximação do parâmetro de relaxação para este sistema linear. Para esta matriz, o valor obtido pela estimativa, foi $\lambda = 1.1541$.

4 Algoritmo Paralelo

No caso em que as projeções são feitas de maneira exata, é fácil ver que a projeção ótima a ser selecionada é a que maximiza $\|x^{k+1} - x^k\|_2$. Infelizmente o custo do cálculo de cada uma

ISSN 2237-7166

dessas normas consome $O(n)$ operações, mais o custo do cálculo da projeção para gerar x^{k+1} que também é $O(n)$. No entanto, este processo pode ser facilmente paralelizado.

Nossa ideia foi selecionar l linhas de forma arbitrária da matriz A e calcular através de processamento paralelo valores para x^{k+1} e as normas $\|x^{k+1} - x^k\|_2$ baseando-se no algoritmo de Eldar e Needell [4], mas sem considerar a redução de dimensionalidade. Para isto, utilizamos a Plataforma de Computação Paralela CUDA e para cada uma das l linhas selecionadas são calculados os valores de x^{k+1} e $\|x^{k+1} - x^k\|_2$ em diferentes *threads* da GPU.

A Figura 3 mostra o resultado do tempo de execução do algoritmo para diferentes quantidades de threads e valores de λ. Essa simulação feita em uma máquina Intel Core i7 X 990 @ 3.47GHz, 250GB RAM e GPU Tesla C2050/C2070, mostra que existe um gargalo ao se utilizar uma quantidade maior que 120 *threads* e, aumentando este número, diminuímos a velocidade de convergência.

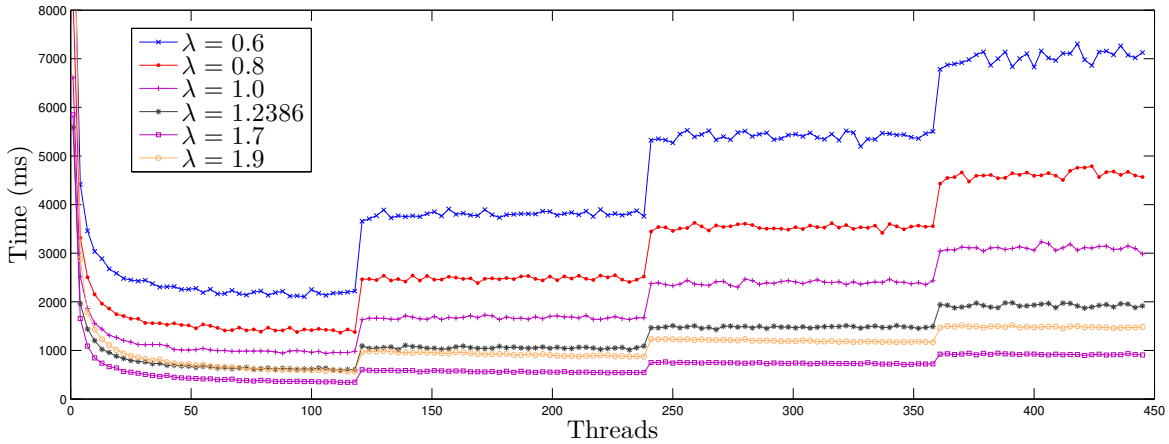


Figura 3: Tempo gasto para convergência do algoritmo Kaczmarz Paralelo com critério de parada $\|x^k - x^*\| < 10^{-2}$ executado na máquina descrita no texto.

A matriz $A \in \mathbb{R}^{m\times n}$ usada para fazer a simulação apresentada na Figura 3 possui dimensões $m = 5000$, $n = 500$ e os valores de seus elementos foram gerados através de uma função geradora de números pseudo-aleatórios uniformemente distribuídos dentro do intervalo $[0, 1]$. Minimizando (9) em função de λ, obtemos o parâmetro de relaxação $\lambda = 1.2386$.

5 Considerações Finais

Apesar do trabalho ainda estar em andamento, tanto a estimativa quanto a implementação dos algoritmos paralelos mostraram-se bastante promissores. Nas várias simulações produzidas, o algoritmo aleatório de Strohmer sempre converge mais rapidamente com o valor de parâmetro escolhido após minimizar (9), isto é, caso comparemos com o mesmo algoritmo utilizando valor de parâmetro $\lambda = 1$.

Porém, notamos que em nossas implementações paralelas, o parâmetro escolhido por (9) ainda fica muito longe dos valores obtidos experimentalmente. Atribuímos este comportamento ao fato de calcularmos a projeção que mais anda em relação a solução do sistema através de $\|x^{k+1} - x^*\|_2$, não respeitamos a esperança da seleção das linhas calculada pela Inequação (8).

Nossas implementações paralelas ainda não contam com a redução de dimensionalidade baseada no lema de Johnson–Lindenstrauss, mas isso será feito em um futuro próximo.

Referências

[1] R. L. Burden and D. J. Faires. *Numerical Analysis*. Brooks Cole, 7 edition, Dec. 2000.

[2] Y. Censor, P. Eggermont, and D. Gordon. Strong underrelaxation in kaczmarz's method for inconsistent systems. *Numerische Mathematik*, 41:83–92, 1983.

ISSN 2237-7166

[3] F. Deutsch and H. Hundal. The rate of convergence for the method of alternating projections, ii. *Journal of Mathematical Analysis and Applications*, 205(2):381 – 405, 1997.

[4] Y. C. Eldar and D. Needell. Acceleration of randomized kaczmarz method via the johnson—lindenstrauss lemma. *Numer. Algorithms*, 58(2):163–177, Oct. 2011.

[5] H. G. Feichtinger, C. Cenker, M. Mayer, H. Steier, and T. Strohmer. New variants of the pocs method using affine subspaces of finite codimension with applications to irregular sampling. pages 299–310, 1992.

[6] H. G. Feichtinger and K. Gröchenig. Theory and practice of irregular sampling. In J. Benedetto and M. Frazier, editors, *Wavelets: Mathematics and Applications*, Studies in Advanced Mathematics, pages 305–363. NuHAG;Classical, CRC Press, 1994.

[7] N. B. Franco. *Cálculo Numérico*. São Paulo: Pearson Prentice Hall, 1 edition, Oct. 2006.

[8] A. Galántai. On the rate of convergence of the alternating projection method in finite dimensional spaces. *Journal of Mathematical Analysis and Applications*, 310(1):30 – 44, 2005.

[9] G. H. Golub and C. F. van Van Loan. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*. The Johns Hopkins University Press, 3rd edition, Oct. 1996.

[10] K. Gröchenig. Irregular sampling, toeplitz matrices, and the approximation of entire functions of exponential type. *Math. Comp*, 68:749–765, 1999.

[11] M. Hanke and W. Niethammer. On the acceleration of kaczmarz's method for inconsistent linear systems. *Linear Algebra and its Applications*, 130(0):83 – 98, 1990.

[12] G. Herman and L. Meyer. Algebraic reconstruction techniques can be made computationally efficient [positron emission tomography application]. *Medical Imaging, IEEE Transactions on*, 12(3):600–609, sep 1993.

[13] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres*, 35:355–357, 1937.

[14] F. Natterer. *The mathematics of computerized tomography*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

[15] D. Needell. Randomized kaczmarz solver for noisy linear systems. *BIT Numerical Mathematics*, 50:395–403, 2010.

[16] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edition, 2003.

[17] K. M. Sezan and H. Stark. Applications of convex projection theory to image recovery in tomography and related areas. page 415–462, 1987.

[18] T. Strohmer and R. Vershynin. A randomized solver for linear systems with exponential convergence. In J. Díaz, K. Jansen, J. Rolim, and U. Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 4110 of *Lecture Notes in Computer Science*, pages 499–507. Springer Berlin Heidelberg, 2006.

[19] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.