数理解析研究所講究録
第2001巻 2016年 64-73

深層学習のリッジレット解析にむけた取組み *

Toward Ridgelet Analysis of Deep Learning

早稲田大学　先進理工学研究科 † 園田　翔　村田　昇
Sho Sonoda　Noboru Murata
Faculty of Science and Engineering
Waseda University

概要

深層学習は圧倒的な学習能力を誇る手法として 2010 年頃から注目を集めている。深層学習で用いる深層ネットワークは，従来の浅いニューラルネットの合成写像とみなせる。浅いニューラルネットは適当な条件のもと L^2 空間で稠密なので，関数近似という観点では深層構造は冗長である。本講演では，ニューラルネットを連続化してリッジレット変換とみなす方法を説明し，浅いネットワークに対する最近の結果を紹介したあと，深層構造への展開を検討する。

1　はじめに

深層学習による技術革新が目覚ましい。画像認識 [1] からアーケードゲーム [2] まで，幅広いタスクで深層学習が従来手法を圧倒している。最近では画像を生成する [3, 4] ことまでできるようになった。AlphaGo と呼ばれる囲碁プログラム [5] が，世界最強とも評されるイ・セドル氏を破ったことも記憶に新しい。深層学習の具体例については既に多くの解説記事が出版されているので，そちらを参照されたい [6, 7, 8, 9]。

深層学習とは，二層以上の隠れ層を備えたニューラルネット（深層ネットワーク）を学習させる手法の総称である。深層学習の発達に伴い，深層ネットワークが高度な情報処理能力をもつことが分かってきたが，その機構の理解は発展途上である。例えば，なぜ深

2015 RIMS 共同研究 「ウェーブレット解析と信号処理」
本研究は JSPS 科研費 15J07517 の助成を受けたものです。
〒169-8555 東京都新宿区大久保 3-4-1

層構造かという根本的な問いかけに対しても，情報を階層的に表現するために効率的であるという立場 [10, 11, 12] や，入力の不変性を抽出するために効率的であるという立場 [13, 14]，ほとんどランダムなのではないかという立場 [15, 16]，実は一層に圧縮できるという報告 [17] など，百家争鳴である。

発表者は，ニューラルネットの積分表現理論（リッジレット解析）[18, 19, 20, 21] を通じて深層ネットワークの解析に取り組んでいる。積分表現はニューラルネットを関数解析的に扱うための強力なツールであり，後に双対リッジレット変換という名前がついた。例えば，積分表現を使うとニューラルネットの隠れ層が計算できる。通常，ニューラルネットは非線形最適化問題の局所解という程度の特徴付けしかできないことと較べて，これは著しい強みである。

積分表現理論では，浅いニューラルネットを基底と係数に分けて考えるため，深層ネットワークに対して積分表現を計算する方法は自明ではない。深層ネットワーク向けの積分表現理論については現在論文投稿中のため，本稿では考え方を紹介するのみに止める。本稿の後半では，研究会当日はまだ投稿中であった Sonoda and Murata [21] の内容を解説する。

2　深層学習のリッジレット解析

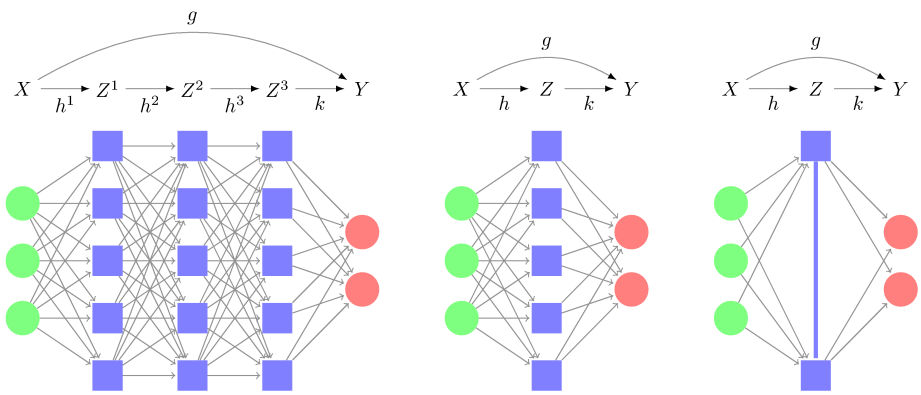

図 1　ニューラルネットの模式図。いずれも 3 入力 2 出力のネットワークを表す。左: 隠れ層を 3 層備えた深層ネットワーク。中央: 従来の（浅い）ニューラルネット。右: 積分表現に相当するニューラルネット。

以降では $X = \mathbb{R}^m, Y = \mathbb{R}^n$ とし，Z^ℓ を適当な次元のユークリッド空間とする。

2.1 浅いニューラルネット

従来用いられてきた「浅い」ニューラルネットとは以下の関数 $g : X \to Y$ である。

$$g(x) = \sum_{j=1}^{J} c_j \eta(a_j \cdot x - b_j), \quad (a_j, b_j, c_j) \in \mathbb{R}^{m \times 1 \times n} \tag{1}$$

ここで $\eta : \mathbb{R} \to \mathbb{R}$ は活性化関数と呼ばれる非線形関数であり，具体的にはガウス関数 $\exp(-z^2/2)$ や双曲線正接関数 $\tanh(z)$，切断ベキ z_+ を用いることが多い。

ニューラルネットの基底関数[*1]

$$h_j(x) := \eta(a_j \cdot x - b_j), \quad j = 1, \cdots, J \tag{2}$$

を $Z = \mathbb{R}^J$ に値をとるベクトル値関数とみなして単に $h(x)$ と書き，$h : X \to Z$ を隠れ層と呼ぶ。係数 c_j が定める線形写像を $k : Z \to Y$ と書くことにすると，浅いニューラルネット (1) は

$$g = k \circ h, \tag{3}$$

と書ける。

2.2 ニューラルネットの積分表現とリッジレット変換

浅いニューラルネットの和を積分に取り替えたものが積分表現である。

$$g(x) = \int_{X \times \mathbb{R}} \mathrm{T}(a, b) \eta(a \cdot x - b) \mathrm{d}a \mathrm{d}b. \tag{4}$$

積分表現は η による T の双対リッジレット変換 $\mathscr{R}_\eta^\dagger \mathrm{T}(x)$ としても知られる。例えばディラックの δ 関数を用いて

$$\mathrm{T}(a, b) = \frac{1}{J} \sum_{j=1}^{J} c_j \delta_{a_j, b_j}(a, b) \tag{5}$$

とすると，元のニューラルネット (1) に帰着する。

[*1] ニューラルネットという対象が先にあるので，厳密に基底ないしフレームあるいはアトムを成す義理はないが，慣例に倣い「基底」と呼ぶ。

関数 $f : X \to Y$ と $\psi : \mathbb{R} \to \mathbb{R}$ に対して，f の ψ によるリッジレット変換を以下で定義する。

$$\mathscr{R}_\psi f(a,b) := \int_X f(x)\overline{\psi(a \cdot x - b)}\mathrm{d}x, \tag{6}$$

η と ψ が許容条件

$$\int_{\mathbb{R}} \frac{\overline{\widehat{\psi}(\zeta)}\widehat{\eta}(\zeta)}{|\zeta|^m}\mathrm{d}\zeta = 1, \tag{7}$$

を満たすとき[*2]，リッジレット変換の再生公式が成り立つ。

$$f(x) = \int_{X\times\mathbb{R}} \mathscr{R}_\psi f(a,b)\eta(a \cdot x - b)\mathrm{d}a\mathrm{d}b. \tag{8}$$

右辺はニューラルネットの積分表現において $\mathrm{T} = \mathscr{R}_\psi f$ としたものなので，これは勝手な関数 f をニューラルネットとして実現する式になっている。従って，右辺を適当な方法で有限和近似すると，関数 f を近似する浅いニューラルネットが得られる [21, 22]。

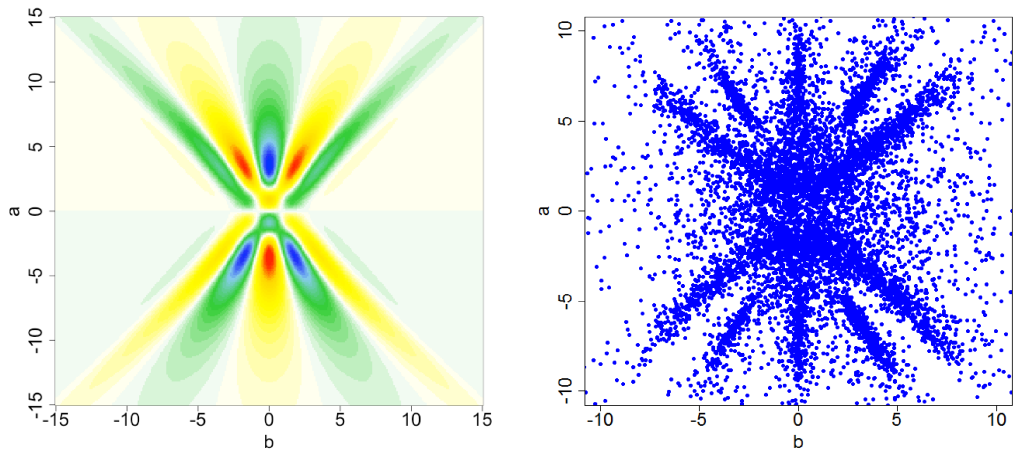

図 2　リッジレット変換と実際の学習結果の比較。左: $f(x) = \sin 2\pi x,\ x \in [-1,1]$ に対するリッジレット変換 $\mathscr{R}_\psi f(a,b)$ の計算例。右: 同 $f(x)$ を 1,000 個のニューラルネットで学習した場合の学習済パラメータの分布。$\mathscr{R}_\psi f(a,b)$ の値が高い位置にあるパラメータ (a,b) は，実際のニューラルネットでも使われやすいことが分かる。

[*2] η が超関数の場合には条件がやや複雑になる。Sonoda and Murata [21, Definifition 5.1] 参照。

2.3 深層ネットワーク

深層ネットワークは，隠れ層を多重化したニューラルネットである。

$$g = k \circ h^L \circ \cdots \circ h^1. \tag{9}$$

ここで各 $h^\ell : Z^{\ell-1} \to Z^\ell$ は (2) の形式とし，$k : Z^L \to Y$ は線形写像とする。

深層ネットワークの構造を調べることは，$h^L \circ \cdots \circ h^1$ の途中にあたる $h^\ell \circ \cdots \circ h^1$ の振る舞いを調べることだが，単純に $\mathscr{R}_\psi f$ を計算するだけでは $h^L \circ \cdots \circ h^1$ の情報しか得られないので，工夫が必要である。

2.4 深層学習のリッジレット解析

深層ネットワークが次のような構造をもつと仮定する。

$$g = \widetilde{k} \circ k^L \circ h^L \circ \cdots \circ k^1 \circ h^1, \tag{10}$$

ただし $k^\ell : Z^\ell \to X$ および $\widetilde{k} : X \to Y$ は線形写像とする。これは，学習対象 f が力学系 $\Phi_\ell : X \to X$ を用いて次のように書けるという仮定と同等である。

$$f = \widetilde{k} \circ \Phi_L \circ \cdots \circ \Phi_1. \tag{11}$$

このとき，各層は $\Phi_\ell = k^\ell \circ h^\ell$ のように対応するので，浅いネットワークに対するリッジレット変換が適用できる。力学系という仮定は，パターン認識の本質がデータの空間 X を超平面で分割する作業であることを考えれば自然である。すなわち，X の中に散乱した入力データを Φ_ℓ によって整理することで，分割時の難易度を下げていると考えるのである。この方針に基づく結果は現在 (2016 年 3 月) 投稿中である。

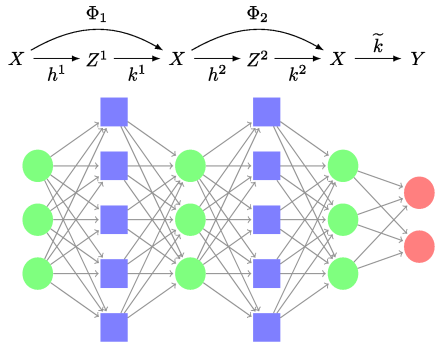

図 3 深層ネットワークが力学系の構造を持つ場合にはリッジレット変換が適用できる。

3　超関数によるリッジレット解析

以降では研究会当日は投稿中であった Sonoda and Murata [21] の内容を紹介する。

深層学習では活性化関数 η として rectified linear unit (ReLU)

$$(z)_+ := \begin{cases} z & z \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

を採用することが多い。ReLU とは切断ベキあるいはランプ関数のことであるが，以降では深層学習での呼び方を踏襲する。従来用いられてきた tanh などと比較して，ReLU は (1) 実装が `if` 文ひとつで済むので計算コストが削減できる (2) 誤差信号が飽和しないので学習が加速する (3) 学習結果がスパースになるというメリットがあり，現在では標準的に用いられている。

従来のリッジレット変換の理論では η として ReLU のような非有界関数は想定していない。このままでは深層ネットワークの積分表現理論を展開するために不便であったので，Sonoda and Murata [21] では η がリゾルキン超関数の場合にもリッジレット解析ができることを示した。つまり，リッジレット変換が定義でき，適当な許容条件のもとで再生公式が成り立つことを示した。ここでリゾルキン超関数とは，シュワルツ超関数であって，多項式を 0 と同一視して得られる超関数のクラスであり，特に ReLU はこのクラスに含まれる。残念ながら η が多項式の場合には再生公式が成り立たないので，ReLU が折れ線であることは本質的である。

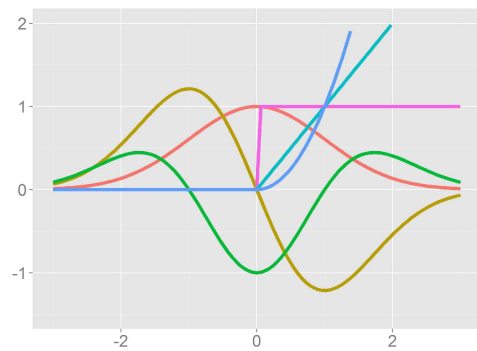

図 4　リゾルキン超関数の例。ガウス関数 $G(z)$ およびその微分 $G'(z), G''(z)$，切断ベキ z_+^0, z_+, z_+^2 が含まれる。

3.1 方針

アイデアは単純である。ReLU を複数組み合わせると有界関数が作れる。例えば,

$$\eta^\star(z) := z_+ - (z-1)_+, \tag{13}$$

とおけば良い。右辺は平行移動作用素 $[\tau^a f](z) := f(z-a)$ を用いて

$$\eta^\star = (\tau^0 - \tau^1)[(\cdot)_+], \tag{14}$$

と書けることに注意する。

$\eta^\star$ は有界関数なので，従来の枠組みの範疇で再生公式が成り立つ。

$$f(x) = \int_{X\times\mathbb{R}} \mathscr{R}_\psi f(a,b)\eta^\star(a\cdot x - b)\mathrm{d}a\mathrm{d}b. \tag{15}$$

ここで，積分表現が b に関して畳込みの形をしていることに着目する。畳込みは線形作用素であり，平行移動作用素と可換なので，形式的に以下が成り立つことが期待される。

$$(15) = \int_{X\times\mathbb{R}} \mathscr{R}_\psi f(a,b)(\tau^0-\tau^1)[(\cdot)_+](a\cdot x-b)\mathrm{d}a\mathrm{d}b \tag{16}$$

$$= \int_{X\times\mathbb{R}} (\tau^0-\tau^1)[\mathscr{R}_\psi f](a,b)(a\cdot x-b)_+\mathrm{d}a\mathrm{d}b. \tag{17}$$

最後の式は ReLU による積分表現なので，積分の存在を示せば証明が完了する。

3.2 概要

Sonoda and Murata [21] では，平行移動を超関数の意味での微分に置き換えて上記の主張を証明した (Theorem 5.6, 5.7, 5.11)。フーリエ変換を参考にして，まず L^1 関数の場合に再生公式が成り立つことを示し，続いて $L^1\cap L^2$ 関数の場合に Parseval の定理が成り立つことを示し，最後にリッジレット変換が有界作用素であることを用いて L^2 関数に対してリッジレット変換が定義できることを示した。

証明の過程ではシュワルツによる関数のクラス分けを元にして，f と ψ のクラスに応じて $\mathscr{R}_\psi f$ がどのクラスに属するかまで踏み込んで調べた (Theorem 4.2)。また，許容的なリッジレット関数 η,ψ の構造を調べ (Theorem 5.4)，許容的関数を構成する方法を導いた (Corollary 5.5)。

Corollary 5.5 η をリゾルキン超関数とし，$\widehat{\eta}$ を η のフーリエ変換とする。ある 0 の近傍 Ω と自然数 k があって，$\zeta^k \cdot \widehat{\eta}(\zeta)$ が Ω 上 C^1 級関数になるとする。急減少関数 ψ_0 として

$$\int_{\mathbb{R}} \zeta^k \overline{\widehat{\psi_0}(\zeta)} \widehat{\eta}(\zeta) \mathrm{d}\zeta = 1, \tag{18}$$

となるものを求めよ。このとき，$\psi := \mathcal{H}^m \psi_0^{(k+m)}$ と η は許容的である。ただし $\mathcal{H}$ はヒルベルト変換である。

この条件は，$\widehat{\eta}$ が分かっている場合には強力なレシピとなる。具体的な計算例は当該論文の §6 で紹介したので，そちらを参照されたい。

4 ラドン変換およびウェーブレット変換との関係

最後に，ニューラルネットの幾何学的な側面について触れる。リッジレット変換はラドン変換とウェーブレット変換の合成変換に分解できる。すなわち，$(u, \alpha, \beta) = (a/|a|, 1/|a|, b/|a|)$ と変数変換したうえで，リッジレット変換を次のように変形できる。

$$\mathscr{R}_\psi f(u, \alpha, \beta)/\alpha = \int_X f(x) \overline{\psi\left(\frac{u \cdot x - \beta}{\alpha}\right) \frac{1}{\alpha}} \mathrm{d}x \tag{19}$$

$$= \int_{\mathbb{R}} \left[\int_{(\mathbb{R}u)^\perp} f(pu + y) \mathrm{d}y \right] \overline{\psi\left(\frac{p - \beta}{\alpha}\right) \frac{1}{\alpha}} \mathrm{d}p \tag{20}$$

$$= \int_{\mathbb{R}} \mathrm{R}f(u, p) \overline{\psi\left(\frac{p - \beta}{\alpha}\right) \frac{1}{\alpha}} \mathrm{d}x, \tag{21}$$

ただし $\mathrm{R}f(u, p)$ は超平面 $x \cdot u = p$ 上で f を積分して得られる f のラドン変換である。

このことから予想される通り，もう一方の積分表現（双対リッジレット変換）は，双対ウェーブレット変換と双対ラドン変換の合成変換になっている。トモグラフィーの分野では，ラドン変換の逆を計算する方法には三通りあることが知られている [23]。すなわち，フーリエスライス定理を用いる方法，逆投影フィルタを計算する方法，逆行列を計算する方法である。Theorem 5.6 ではリッジレット変換のフーリエスライス定理を用いてリッジレット変換をフーリエ変換に帰着し，再生公式を示した。Theorem 5.7 では，リッジレット変換の再生公式がラドン変換の再生公式に帰着することを示して，元の再生公式を証明した。このとき，許容条件はウェーブレット変換と双対ウェーブレット変換が逆投影フィルタに変形できるための条件であることが分かった。

このように，ラドン変換やウェーブレット変換は幾何学的な背景を持つので，ニューラルネットもまた幾何学的に考察できるのである。

参考文献

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, Léon Bottou, and K. Q. Weinberger, editors, *Adv. Neural Inf. Process. Syst. 25*, pp. 1097–1105. Curran Associates, Inc., 2012.

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei a Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, Vol. 518, No. 7540, pp. 529–533, 2015.

[3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In *Int. Conf. Learn. Represent. 2016*, pp. 1–15, 2015.

[4] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A Neural Algorithm of Artistic Style. arXiv:1508.06576, 2015.

[5] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, Vol. 529, No. 7587, pp. 484–489, 2016.

[6] 得居誠也. Deep Learning 技術の今, `http://www.slideshare.net/beam2d/deep-learning20140130`, 2016/3/1, 2014.

[7] 麻生英樹, 安田宗樹, 前田新一, 岡野原大輔, 岡谷貴之, 久保陽太郎, ダヌシカボレガラ. 深層学習. 近代科学社, 2015.

[8] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: a review and new perspectives. *Pattern Anal. Mach. Intell. IEEE Trans.*, Vol. 35, No. 8, pp. 1798–1828, 2013.

[9] Jürgen Schmidhuber. Deep Learning in neural networks: An overview. *Neural Networks*, Vol. 61, pp. 85–117, 2015.

[10] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, Vol. 18, No. 7, pp. 1527–1554, 2006.

[11] Yoshua Bengio. Learning Deep Architectures for AI. *Found. Trends Mach. Learn.*, Vol. 2, No. 1, pp. 1–127, 2009.

[12] Honglak Lee. *Unsupervised Feature Learning via Sparse Hierarchical Representations*. PhD thesis, Stanford University, 2010.

[13] Jake Bouvrie, Lorenzo Rosasco, and Tomaso Poggio. On invariance in hierarchical models. In *Adv. Neural Inf. Process. Syst.*, Vol. 22, pp. 162–170, 2009.

[14] Joan Bruna and Stéphane Mallat. Invariant Scattering Convolution Networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 35, No. 8, pp. 1872–1886, 2013.

[15] Raja Giryes, Guillermo Sapiro, and Alex M Bronstein. Deep Neural Networks with Random Gaussian Weights : A Universal Classification Strategy ? arXiv:1504.08291, 2015.

[16] Andrew M Saxe, Pang Wei Koh, Zhenghao Chen, Maneesh Bhand, Bipin Suresh, and Andrew Y Ng. On Random Weights and Unsupervised Feature Learning. *28th Int. Conf. Mach. Learn.*, pp. 1–9, 2011.

[17] Lj Ba and R Caurana. Do Deep Nets Really Need to be Deep ? In *Adv. Neural Inf. Process. Syst. 27*, pp. 2654–2662, 2014.

[18] Noboru Murata. An Integral representation of functions using three-layered betworks and their approximation bounds. *Neural Networks*, Vol. 9, No. 6, pp. 947–956, aug 1996.

[19] Emmanuel Jean Candès. *Ridgelets: Theory and Applications.* PhD thesis, Stanford University, 1998.

[20] S Kostadinova, S Pilipović, K Saneva, and J Vindas. The ridgelet transform of distributions. *Integr. Transform. Spec. Funct.*, Vol. 25, No. 5, pp. 344–358, 2014.

[21] Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. *Appl. Comput. Harmon. Anal.*, 2015.

[22] Sho Sonoda and Noboru Murata. Sampling hidden parameters from oracle distribution. In *24th Int. Conf. Artif. Neural Networks*, Vol. 8681, pp. 539–546, Hamburg, Germany, 2014. Springer International Publishing.

[23] Frank Natterer. X-ray tomography. In Luis L. Bonilla, editor, *Inverse Probl. Imaging*, Vol. 1943, pp. 17–34. Springer-Verlag Berlin Heidelberg, 2008.