Title	L1ノルムを用いた点列の一連節折れ線近似アルゴリズムの開発に関する研究
Author(s)	角川，肇喜
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1855
Rights	
Description	Supervisor:浅野　哲夫，情報科学研究科，修士

L_1 ノルムを用いた点列の一連節折れ線近似アルゴリズムの開発に関する研究

角川 肇喜 (310068)

北陸先端科学技術大学院大学 情報科学研究科

2005 年 2 月 10 日

キーワード: 回帰分析，折れ線近似問題，直線近似問題，L_1 ノルム，L_∞ ノルム.

近年の計算機の発達は情報の電子化を促進し，様々な仕事を計算機が行う事を可能とした．それと並行して計算機問題の複雑化と大規模化が進み，問題を高速に解くアルゴリズム開発の重要性が増している．特に実験データ等の整理に用いる回帰分析問題においてもデータの大規模化が進み，これらの問題を効率的に解くアルゴリズムの研究は以前から行われてきた．その成果の１つとして，x-y 平面上の n 個のデータ点からなる点列の直線近似を L_1 ノルムを用いて求める問題については，線形時間のアルゴリズムが既に開発されている．しかし，その発展問題である点列を近似する x 単調な折れ線近似を求める問題については，現時点において精度と実用性を共に満たすアルゴリズムは開発されていない．そこで本研究では一般的な折れ線近似問題の基礎となる屈折点が一つの折れ線近似問題を，L_1 ノルムを用いて効率的に精度良く求める実用的なアルゴリズムの開発を行う．

平面上に与えられた点列の直線近似を求める問題の解法として，最も良く知られているのが最小二乗法である．この手法にみられるように点列を直線等で近似する際には，その近似精度の指標として点列と近似直線の垂直距離を誤差基準（ノルム）として用いるのが一般的である．本研究では，特に近似直線と点列の垂直距離の総和出表される L_1 ノルムに注目し，それを用いた一連節折れ線近似アルゴリズムの開発を行う．他に代表的な誤差基準として最小二乗法の L_2 ノルム，近似直線と点列の垂直距離の最大値で表される L_∞ ノルム等がある．L_1 ノルムを用いた直線近似はそれら誤差基準に比べ，外れ値やノイズの影響を受けにくいという特徴をもつ一方，問題としては難しくなっている．

折れ線近似アルゴリズムの現状は，折れ線の屈折点数が１以上の定数個の場合，L_∞ ノルムに関しては有効な最適アルゴリズムが既に存在する．また，L_∞ ノルムの最大値を先に決める事により線形時間で屈折点数最小の折れ線を求める線形時間アルゴリズムも存在する．しかし，他の誤差基準に関してはいまだ有効なアルゴリズムは開発されていない．L_1 ノルムについてみると屈折点数が１つの場合については現時点で最も計算時間の良い最適アルゴリズムとして，$O(n^{1.3})$ のものが存在する．しかし，この方法は動的かつ複雑

なデータ構造を用いており実用には適さないという問題点がある．本研究では，特に実用性を重視し，将来的には屈折点数を一般化することを目的とした，L_1 ノルムを用いた屈折点１つの折れ線近似アルゴリズムの開発を行った．

L_1 ノルムを用いた一連節折れ線近似問題の最も単純なアルゴリズムは全探索になる．平面上に与えられた n 点それぞれで点列を左右に分割し，その分割ごとに左右で線形時間の直線近似アルゴリズムを用いて二本の直線を求め，それにより折れ線を構成する．そうして求めた n 個の折れ線の内，最も L_1 ノルムが小さなものを点列に対する折れ線として出力する．この方法の計算時間は $O(n^2)$ となる．本研究ではこの方法を基に改善を試みる事を考える．改善の方針としては二通りの方法が考えられる．一方は，点列の分割点候補の絞り込みを行うことにより，少ない数の分割点候補に対してのみ折れ線近似を行い計算時間を減らす方法である．もう一方は，分割点候補の数は減らさず，各分割点における，直線近似のアルゴリズムを改善する事により計算時間を減らす方法である．本研究では特に前者の考えに基づいたアルゴリズムの開発を行った．また，後者の考えに基づき計算時間を短縮したのが，先に述べた $O(n^{1.3})$ のアルゴリズムである．

アルゴリズム改善の最大の問題点は，屈折点候補をどのようにして絞り込むかという点にある．ここでは，この問題の解決に２通りの方法を提案する．１つ目は，L_∞ 折れ線近似の屈折点を L_1 折れ線近似の屈折点候補とし，それら候補点を屈折点とした折れ線の内，最も良いものを結果として出力する方法である．２つ目の方法では，L_1 ノルムによる近似直線は平面を上下に，それぞれ同数の点を含む２つの半平面に分割するという特性を利用して，折れ線近似の屈折点候補を求めている．両者共に上の様にして求めた１つの点のみでなく，その周辺の数個の点を候補点として利用する．前者の長所は L_1 ノルムと，L_∞ ノルムの互いの長所が，互いの短所を補っている点にある．後者の長所は前者に比べて，より最適な屈折点に近い屈折点候補が得られる点にある．短所としては，両者共に最終的に最適な折れ線近似が得られる保証が無いという点と，後者においては屈折点数の一般化に対応しにくいという点が挙げられる．また，本研究の対象は一連節の折れ線近似アルゴリズムの開発なので，それをするに相応しい形状の点列に対してのみ，この手法を用いることを考える．この判定には L_∞ ノルムを用いた折れ線近似の結果を用いる．いくつかの条件を設けて有効な屈折点数が一つであると判定された場合のみ折れ線近似を行うこととした．

以上の考えに基づき，線形時間のアルゴリズムを実装し，様々な点列に対して実行することにより，最適な折れ線近似と各手法による結果を比較し評価を行った．その結果，点列の一連節折れ線近似については L_∞ 折れ線近似の結果を利用する方法に比べ，L_1 直線近似の結果より屈折点候補を絞る方法の方が，より最適に近い折れ線近似が得られることが分かった．