ARTICLE IN PRESS
YACHA:1150

Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

The mixed Lipschitz space and its dual for tree metrics

William Leeb [a,b,*]

[a] Department of Mathematics, Yale University, New Haven, CT 06511, United States
[b] PACM, Princeton University, Princeton, NJ 08544, United States

A R T I C L E I N F O

Article history:
Received 25 January 2015
Received in revised form 5 November 2015
Accepted 23 June 2016
Available online xxxx
Communicated by Dominique Picard

MSC:
68W01
05C05
26A16
54E35
46M05

Keywords:
Tree metric
Lipschitz
Mixed Lipschitz
Dual space
Earth Mover's Distance
EMD
Martingale difference

A B S T R A C T

This paper develops a theory of harmonic analysis on spaces with tree metrics, extending previous work in this direction by Gavish, Nadler and Coifman (2010) [30] and Gavish and Coifman (2011, 2012) [28,29]. We show how a natural system of martingales and martingale differences induced by a partition tree leads to simple and effective characterizations of the Lipschitz norm and its dual for functions on a single tree metric space. The restrictions we place on the tree metrics are far more general than those considered in previous work. As the dual norm is equal to the Earth Mover's Distance (EMD) between two probability distributions, we recover a simple formula for EMD with respect to tree distances presented by Charikar (2002) [36].

We also consider the situation where an arbitrary metric is approximated by the average of a family of dominating tree metrics. We show that the Lipschitz norm and its dual for the tree metrics can be combined to yield an approximation to the corresponding norms for the underlying metric.

The main contributions of this paper, however, are the generalizations of the aforementioned results to the setting of the product of two or more tree metric spaces. For functions on a product space, the notion of regularity we consider is not the Lipschitz condition, but rather the mixed Lipschitz condition that controls the size of a function's mixed difference quotient. This condition is extremely natural for datasets that can be described as a product of metric spaces, such as word-document databases. We develop effective formulas for norms equivalent to the mixed Lipschitz norm and its dual, and extend our results on combining pairs of trees.

1. Introduction

This paper develops a theory of harmonic analysis on spaces endowed with tree metrics, which are distances that arise naturally throughout pure and applied mathematics. We are concerned primarily with spaces of Lipschitz and mixed Lipschitz functions and their duals, and in particular, simple and computable characterizations of the norms on these spaces.

* Correspondence to: PACM, Princeton University, Princeton, NJ 08544, United States.
E-mail address: wleeb@math.princeton.edu.

http://dx.doi.org/10.1016/j.acha.2016.06.008
1063-5203/

Please cite this article in press as: W. Leeb, The mixed Lipschitz space and its dual for tree metrics, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.06.008

1.1. The Lipschitz space, the dual space and Earth Mover's Distance

Given a metric space (X, d), a natural way of measuring the variation of a function f defined on X is its Lipschitz norm, defined by

$$\sup_{x \neq y} \frac{f(x) - f(y)}{d(x, y)}. \tag{1}$$

Oftentimes, it is convenient to define the Lipschitz norm to be the sum or maximum of (1) and $\|f\|_\infty$; we will consider both versions in this paper, though for the purposes of this introductory section we will restrict our attention to (1). If f is a differentiable function on $\mathbb{R}$, the Lipschitz norm (1) is equal to $\|f'\|_\infty$, the supremum of f's derivative. Expression (1), however, is defined for non-differentiable functions and makes sense in the abstract setting of any metric space.

The space of Lipschitz functions defined on a metric space arises naturally in many areas of machine learning and statistics. For example, standard models in non-parametric statistics posit that unknown signals lie in a Hölder space (where the underlying metric space is $\mathbb{R}$ and the distance is defined as $d(x, y) = |x-y|^\alpha$ for some $0 < \alpha < 1$) or a more general regularity class [1,2]. Extrapolating a function value to new points, or inferring its values from noisy samples, can only be achieved if some kind of regularity on the function is assumed, the Lipschitz condition being a natural kind of regularity.

In the Euclidean setting of classical analysis, where we consider the space $\mathbb{R}$ equipped with the distance $d(x, y) = |x - y|^\alpha$ for some $0 < \alpha < 1$, (1) (which is then referred to as the Hölder norm of f) can be shown to be equivalent in size to a number of other expressions that look at differences of averages of f over different scales. For example, if we take a sufficiently nice wavelet basis $\{\psi_{j,k}\}$ of $\mathbb{R}^n$ (where $j \in \mathbb{Z}$ indexes the dyadic scale 2^j and $k \in \mathbb{Z}$ the location), then the expression

$$\sup_{j,k} 2^{-j(\alpha+1/2)} |\langle f, \psi_{j,k} \rangle| \tag{2}$$

is equivalent in size to (1), which is to say that the ratio of the two quantities is bounded above and below by finite constants not depending on f [3]. The wavelet coefficients $\langle f, \psi_{j,k} \rangle$ can be thought of as measuring f's variation across scales.

In a discrete setting, where we only have f sampled on a grid of k points in $\mathbb{R}^n$, computing the Lipschitz norm (1) directly would require $O(k^2)$ operations, as all pairs of points need to be accounted for. However, using the fast wavelet transform (see, for instance, [4]), the expression (2) can be computed with only $O(k)$ operations. In addition to their computational tractability, the simple characterization of the Hölder norm of a function via its wavelet coefficients gives rise to efficient statistical procedures for signal recovery in the nonparametric setting; see [5].

Also of interest is the space dual to Lipschitz. Given any normed space $(\mathcal{X}, \|\cdot\|)$, one defines its dual space as the collection of linear functionals on $\mathcal{X}$, equipped with the norm

$$\|T\|_* = \sup_{f \in \mathcal{X}: \|f\| \leq 1} \langle f, T \rangle. \tag{3}$$

When $\mathcal{X}$ is the space of Lipschitz functions over a metric space (X, d), the dual norm (3) has another interpretation, described by the Kantorovich–Rubinstein Theorem [6,7]. If μ and ν are two probability measures over X, we define their Earth Mover's Distance (EMD) to be

$$\mathrm{EMD}(\mu, \nu) = \inf_\pi \int\limits_{\Omega \times \Omega} d(x, y) d\pi(x, y). \tag{4}$$