Approximate earth mover's distance in linear time

Sameer Shirdhonkar and David W. Jacobs
{ssameer,djacobs}@umiacs.umd.edu http://www.umiacs.umd.edu/~ssameer/

Comparing probability distributions

Applications: Hypothesis testing, Feature matching in computer vision.

Methods: χ^2 test, simple L_2 norm between densities, EMD/KR metric.

Advantages of EMD: Error in sample values causes shift in estimated distributions. Unlike others, this is not heavily penalized by EMD.

Disadvantage of EMD: $O(n^3 \log n)$ computation time. So fast approximation needed !

KR metric / EMD

P_1, P_2	Probability distributions on a compact metric space S
p_1, p_2	Probability densities
$p := p_1 - p_2$	Difference density
$c(x,y) = \|x-y\|^s$	Cost function
$q(x,y)$	Joint density with p_1 and p_2 as marginals
$f(x)$	Potential function

$$\text{EMD} := \inf_q \int \|x-y\|^s \, q(x,y) dx dy$$

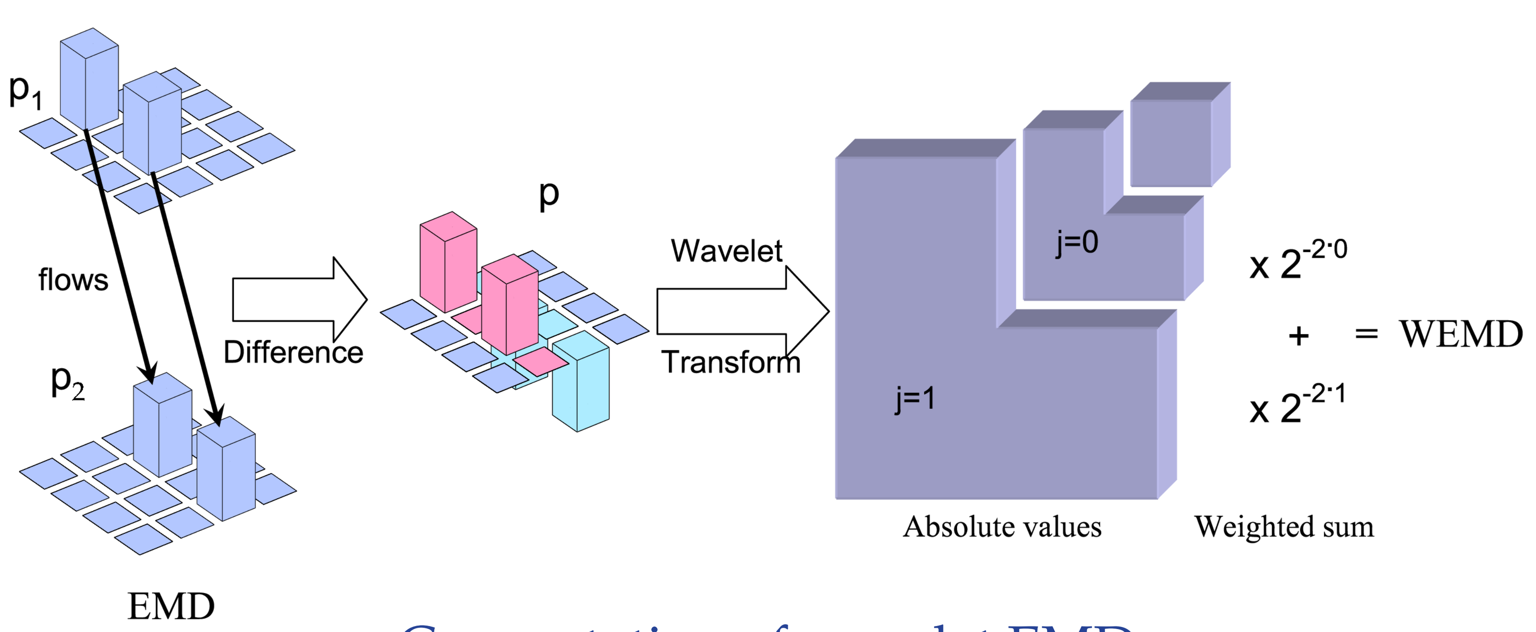
$$\text{s.t.} \int q(u,y)dy - \int q(x,u)dx = p(u)$$

Theorem 1 (Kantorovich-Rubinstein [2]).

$$\textit{Dual EMD} := \sup_f \int f(x)p(x)dx \longrightarrow \textbf{\textit{Inner Product}}$$

$$\textit{s.t.} \, f(x) - f(y) \le \|x-y\|^s \longrightarrow \textbf{\textit{Hölder continuity of order s}}$$

Hölder continuity : $C_H(f) := \sup_{x \neq y} \frac{|f(x)-f(y)|}{\|x-y\|^s}$ exists and is finite.



Computation of wavelet EMD

Wavelet approximation

Theorem 2 ([3]). *$f \in L^1_{loc}(\mathbb{R}^n)$, belongs to $C^s(\mathbb{R}^n)$ if and only if, in a wavelet decomposition of regularity $r \ge 1 > s$ there exist constants C_0 and C_1 such that,*

$$\textit{Approx. coeffs.:} \quad |f_k| \le C_0, \quad k \in \mathbb{Z}^n \quad \textit{and}$$
$$\textit{Detail coeffs.:} \quad |f_\lambda| \le C_1 2^{-j(n/2+s)}, \quad \lambda \in \Lambda_j, \quad j \ge 0 \tag{1}$$

Lemma 1. *If $0 < s < 1$ and (1) holds, then $f \in C^s(\mathbb{R}^n)$ with $C_H(f) < C$ such that*

$$a_{12}(\psi; s)C_1 \le C \le a_{21}(\psi; s)C_0 + a_{22}(\psi; s)C_1 \tag{2}$$

For discrete distributions, if we change the definition of $C_H(f)$ to $C_H(f) := \sup_{|x-y|\ge 1} \frac{|f(x)-f(y)|}{\|x-y\|^s}$ then the same holds for $s = 1$ as well.

Dual EMD in wavelet domain:

$$\text{Maximize } \mathbf{p}^T\mathbf{f} = \textstyle\sum_k p_k f_k + \sum_\lambda p_\lambda f_\lambda$$
$$\text{subject to} \quad |f_k| \le C_0 \quad \text{and} \quad |f_\lambda| \le C_1 2^{-j(s+n/2)} \tag{3}$$

Theorem 3 (Main result). *(p_k, p_λ) are wavelet coefficients of the difference density p. Then for any constants $C_0 \ge 0$ and $C_1 > 0$,*

$$\textit{WEMD} := C_0 \sum_k |p_k| + C_1 \sum_\lambda 2^{-j(s+n/2)} |p_\lambda| \tag{4}$$

is an equivalent metric to the KR metric; i.e. there exist constants $C_L > 0$ and $C_U > 0$ such that

$$C_L \cdot \textit{WEMD} \le \textit{EMD} \le C_U \cdot \textit{WEMD} \tag{5}$$

For discrete distributions, the same result holds for $s = 1$ as well.

$$\text{WEMD} = \sum_\lambda |p_\lambda| 2^{-j(s+n/2)} \tag{6}$$

(Since C_0, C_1 are arbitrary.) This is an $O(n)$ time computation !

[1] P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *3rd International Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[2] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems I: Theory*. Springer, 1998.

[3] Y. Meyer. *Wavelets and Operators, Vol 1*. Cambridge university press, 1992.

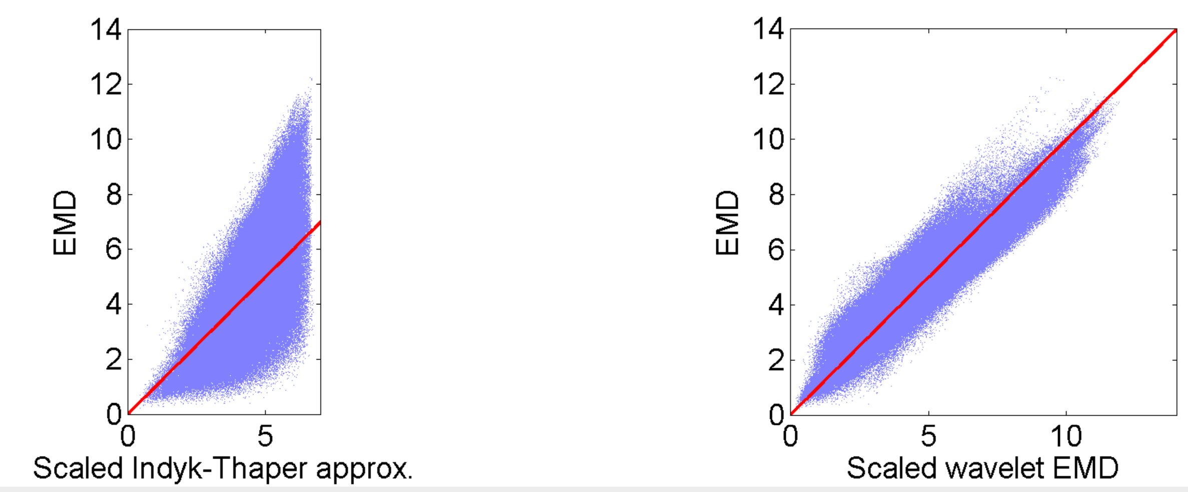
Experiments

Daubechies	C_U/C_L	Daub. symmetric	C_U/C_L
db3	6.33	sym3	6.33
db4	7.29	sym4	4.64
db5	9.92	sym5	6.01
db6	12.59	sym6	5.58
Coiflets	**C_U/C_L**	**Ojanen**	**C_U/C_L**
coif1	4.38	oj8	7.46
coif2	4.75	oj10	10.56
coif3	5.85	oj12	13.79

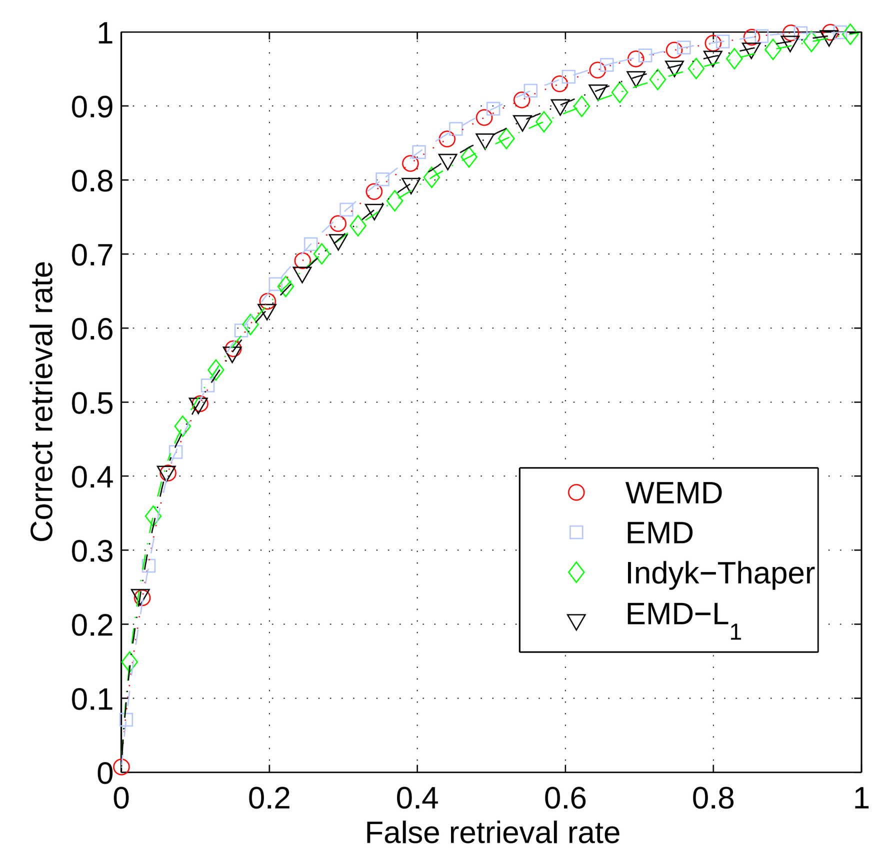
Theoretical (loose) estimates for MAX error for 1D wavelets.

Method	Bounds ratio	Normalized RMS error	Preproc. time	Compare time
EMD	–	–	0.92 s	63 ms
Wavelet EMD	7.03	18%	2.35 s	0.11 ms
Indyk-Thaper	11.00	43%	0.51 s	22 ms

Error and time requirements for 16x16x16 colour histograms



WEMD (coif3) has less error than Indyk and Thaper's [1] method.



Colour histograms for content based image retrieval: wavelet EMD performance compared to other EMD methods