8.5 Wavelets and VC-dim

Wednesday, March 4, 2020 8:18 PM

Sufficiency conditions for orthogonal wavelets

Lemma 11.7 If $b_k = (-1)^k c_{d-1-k}$, then $\int_{-\infty}^{\infty} \phi(x)\psi(2^j x - \ell)\,dx = 0 \quad \forall j, \ell.$

proof.

$$\int_{-\infty}^{\infty} \phi(x)\psi(x-k)\,dx = \int_{-\infty}^{\infty} \sum_{i=0}^{d-1} c_i \phi(2x-i) \sum_{j=0}^{d-1} b_j \phi(2x-2k-j)\,dx$$

$$= \sum_{i=0}^{d-1}\sum_{j=0}^{d-1} c_i b_j \int_{-\infty}^{\infty} \phi(2x-i)\,\phi(2x-2k-j)\,dx$$

$$= \frac{1}{2}\sum_{i=0}^{d-1}\sum_{j=0}^{d-1} c_i b_j \int_{-\infty}^{\infty} \phi(y)\,\phi(y+i-2k-j)\,dy$$

$$= \frac{1}{2}\sum_{i=0}^{d-1}\sum_{j=0}^{d-1} (-1)^j c_i c_{d-1-j}\, \delta(i-2k-j) \qquad (i = 2k+j)$$

$$= \frac{1}{2}\sum_{j=0}^{d-1} (-1)^j c_{2k+j}\, c_{d-1-j}$$

(because d is even) $$= \frac{1}{2}\left[c_{2k}c_{d-1} - c_{2k+1}c_{d-2} + \cdots + c_{d-2}c_{2k-1} - c_{d-1}c_{2k}\right] = 0.$$

Thus, $\int_{-\infty}^{\infty} \phi(2^j x - m)\,\psi(2^j x - k)\,dx = 0 \quad \forall\, m, k$ by substitution.

Note that $\phi(x)$ can be expressed as a linear combination of $\phi(2^j x - m)$ for some m

Thus,

$$\int_{-\infty}^{\infty} \phi(x)\psi(2^j x - k)\,dx = \int_{-\infty}^{\infty} \sum_m \phi(2^j x - m)\,\psi(2^j x - k)\,dx$$

$$= \sum_m \int_{-\infty}^{\infty} \phi(2^j x - m)\,\psi(2^j x - k)\,dx = 0.$$

Lemma 11.8 If $b_k = (-1)^k c_{d-1-k}$, then

$$\int_{-\infty}^{\infty} \frac{1}{2^j}\psi(2^j x - k)\cdot \frac{1}{2^\ell}\psi(2^\ell x - m)\,dx = \delta(j-\ell)\,\delta(k-m).$$

proof.

$$\int_{-\infty}^{\infty} \psi(x)\psi(x-k)\,dx \approx \sum_{i=0}^{d-1}\sum_{j=0}^{d-1} b_i b_j\, \delta(i-2k-j) = \sum_{i=0}^{d-1} b_i b_{i-2k} \qquad \text{(same scale)}$$

$$= \sum_{i=0}^{d-1} (-1)^i c_{d-1-i}\cdot(-1)^{i-2k} c_{d-1-i+2k} = \sum_{i=0}^{d-1} c_{d-1-i}\, c_{d-1-i+2k}$$

Let $j = d-1-i$

$$= \sum^{d-1} \cdots\; 2\delta(k)$$

Let $j = d-1-i$ $\longrightarrow$ $$= \sum_{j=0}^{d-1} c_j c_{j+2k} = 2\delta(k)$$ $\uparrow$ Lemma 11.2

$$\Rightarrow \int_{-\infty}^{\infty} \psi(2^{\ell}x - k)\,\psi(2^{\ell}x - m)\,dx = 2^{2\ell}\delta(k-m).$$

Also, $$\int_{-\infty}^{\infty} \psi(x)\,\psi(2^{\ell}x - k)\,dx = \int_{-\infty}^{\infty} \sum_{i=0}^{d-1} b_i \phi(2x-i)\,\psi(2^{\ell}x - k)\,dx$$

$$= \sum_{i=0}^{d-1} b_i \underbrace{\int_{-\infty}^{\infty} \phi(2x-i)\,\psi(2^{\ell}x-k)\,dx}_{=0 \text{ so long as } \ell \geq 1.} = 0$$

$\square$

Thus, we have both necessary & sufficient conditions to generate an orthonormal basis set of wavelets

$$\phi(x) = \sum_{k=0}^{d-1} c_k \phi(2x-k) \qquad \psi(x) = \sum_{k=0}^{d-1} b_k \phi(2x-k)$$

$$\sum_{k=0}^{d-1} c_k = 2 \quad \Big\} \text{ 1 linear relation} \qquad \underbrace{b_k = (-1)^k c_{d-1-k}}_{\text{no degrees of freedom}}$$

d is even.

$$\underbrace{\sum_{k=0}^{d-1} c_i c_{i-2k} = 2\delta(k) \text{ for } 1 \leq k \leq \frac{d-1}{2}}_{\frac{d}{2} \text{ quadratic equations}}$$ (because otherwise a trivial $0=0$)

Thus, we start with d degrees of freedom in the c_k's, but they have to satisfy $\frac{d}{2}+1$ relations.

$\Rightarrow$ We have $\frac{d}{2}-1$ degrees of freedom to design our wavelet system for properties of interest (e.g. smoothness).

Wavelet Transform

If $f(x) = \sum_{k=0}^{\infty} a_{jk}\phi_{jk}(x)$ where $\phi_{jk}(x) = \phi(2^j x - k)$

If $f(x) = \sum_{k=0} a_{jk} \phi_{jk}(x)$ where $\phi_{jk}(x)$ ...

then $a_{jk} = \int_{-\infty}^{\infty} f(x)\, \phi_{jk}(x)$ by orthogonality.

$$= \int_{-\infty}^{\infty} f(x) \sum_{m=0}^{d-1} c_m\, \phi_{j+1,\, 2k+m}(x)\, dx$$

$$= \sum_{m=0}^{d-1} c_m \int_{-\infty}^{\infty} f(x)\, \phi_{j+1,\, 2k+m}(x)\, dx$$

$$= \sum_{m=0}^{d-1} c_m\, a_{j+1,\, 2k+m}.$$

Let $n = 2k+m$, so $m = n-2k$.

$$\Rightarrow \quad a_{jk} = \sum_{n=2k}^{d-1} c_{n-2k}\, a_{j+1,\, n}$$

This gives us a formula for moving up the tree to compute coefficients of the scale function from high resolution "samples". But, of course, using that we can compute coefficients in the wavelet basis.

Statistical classification

Define: A set system $(X, \mathcal{H})$ consists of a set X and a a class $\mathcal{H}$ of subsets of X.

Ex. X = set of all possible emails

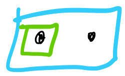
$\mathcal{H} = \{h_0, h_1, h_2, \ldots\}$ where $h_0 \subseteq X$ is the set of all spam emails
$h_1 \subseteq X$ is the set of marketing emails
$\vdots$

Define: A set system $(X, \mathcal{H})$ shatters a set A if each subset of A can be expressed as $A \cap h$ for some $h \in \mathcal{H}$.

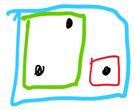
Define: The Vapnik-Chervonenkis (VC) dimension of $\mathcal{H}$ is the size of the largest set shattered by $\mathcal{H}$.

Ex. Let $X = \mathbb{R}^2$ and $\mathcal{H} = \{\text{axis-parallel rectangles}\}$

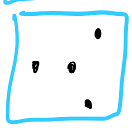
Trivial to shatter.



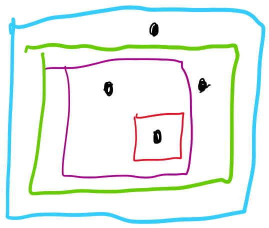
Almost trivial to shatter



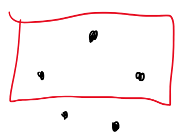
Again, we can always shatter 3 points



We can't get the subset of size 3 containing just the outer points



But there does exist a subset we can shatter. We can get any subset of these 4 points with axis-parallel rectangles.



Can you shatter a set of 5 points?

Consider a rectangle containing all points, and shrink to try to get all sets of 4.

$\Rightarrow$ $VC(\{\text{axis-parallel rectangles}\}) = 4$.

Ex. $X = \mathbb{R}$, $\mathcal{H} = \{[a,b] \mid a,b \in \mathbb{R}\}$. $VC(\mathcal{H}) = 2$.

Ex. $X = \mathbb{R}$ $\mathcal{H} = \{[a,b], [c,d] \mid a,b,c,d \in \mathbb{R}\}$ $VC(\mathcal{H}) = 4$.

Ex. $X = \mathbb{R}$ $\mathcal{H} = \{A \subseteq \mathbb{R} \mid |A| < \infty\}$ $VC(\mathcal{H}) = \infty$

Ex. $X = \mathbb{R}$ $\mathcal{H} = \{\text{convex polygons}\}$ $VC(\mathcal{H}) = \infty$ (via circle)

Ex. $X = \mathbb{R}^d$ $\mathcal{H} = \{\vec{x} \mid \vec{w}^T\vec{x} \geq t\}$ $VC(\mathcal{H}) = d+1$

affine half-spaces

Lemma: $VC(\mathcal{H}) \geq d+1$, where $\mathcal{H} = \{\vec{x} \mid \vec{w}^T\vec{x} \geq t\}$.

Let $S = \{0, e_1, e_2, \dots, e_d\}$ where $e_i = (0, 0, \dots, 0, \underset{\text{pos } i}{1}, 0, \dots, 0)^T$

Let $A \subseteq S$ be a subset. WLOG, assume $0 \in A$.

Let $\vec{w} = (1, 1, \dots, 1)^T - \sum_{a \in A} A$.

Then $A \subseteq \{\vec{x} \mid \vec{w}^T\vec{x} \leq 0\}$ and $S - A \subseteq \{\vec{x} \mid \vec{w}^T\vec{x} > 0\}$.

Let $\vec{w} = (1,1,\dots,1)^T - \sum_{a \in A} \dots$

Then $A \subseteq \{\vec{x} \mid \vec{w}^T\vec{x} \leq 0\}$ and $S - A \subseteq \{\vec{x} \mid \vec{w}^T\vec{x} > 0\}$.

So $\mathcal{H}$ shatters a $d+1$ pt set $\Rightarrow VC(\mathcal{H}) \geq d+1$. $\boxed{\diagup}$

Thm 5.9 (Radon): Any set $S \subseteq \mathbb{R}^d$ with $|S| \geq d+2$ can be partitioned into two disjoint subsets A and B such that $\text{convex}(A) \cap \text{convex}(B) \neq \emptyset$.

proof. WLOG, assume $|S| = d+2$, with $S = \{\vec{a}_1, \dots, \vec{a}_{d+2}\}$

Let $A = [\vec{a}_1 \; \vec{a}_2 \; \vec{a}_3 \; \dots \; \vec{a}_{d+2}] \in \mathbb{R}^{d \times (d+2)}$.

Let $B = \begin{bmatrix} A \\ \vec{1}^T \end{bmatrix}$. $\text{rank}(B) \leq d+1$, so columns are lin. dep. Let $B = [\vec{b}_1 \; \cdots \; \vec{b}_{d+2}]$.

Let $\vec{x} = (x_1, x_2, \dots, x_{d+2})^T$ s.t. $B\vec{x} = 0$

WLOG, say $x_1, \dots, x_s \geq 0$ and $x_{s+1}, \dots, x_{d+2} < 0$.

Normalize $\vec{x}$ s.t. $\sum_{i=1}^{s} |x_i| = 1$.

Then
$$\sum_{i=1}^{s} |x_i| \vec{b}_i = \sum_{i=s+1}^{d+2} |x_i| \vec{b}_i$$

$$\Rightarrow \sum_{i=1}^{s} |x_i| \vec{a}_i = \sum_{i=s+1}^{d+2} |x_i| \vec{a}_i \quad \text{and} \quad \sum_{i=1}^{s} |x_i| = \sum_{i=s+1}^{d+2} |x_i| = 1$$

Thus, both sides are convex combinations of disjoint columns of A.

The convex hulls of the two sets of corresponding pts intersect. $\boxed{\diagup}$

But then it is impossible to have a linear separator of these two sets, so half-planes cannot shatter a $(d+2)$-pt set.

$$\Rightarrow VC(\mathcal{H}) < d+2$$

$$\Rightarrow VC(\mathcal{H}) = d+1.$$

Ex. $X = \mathbb{R}^d$, $\mathcal{H} = \{\vec{x} \mid \|\vec{x} - \vec{v}\| \leq r\}$ spheres

Ex. $X = \mathbb{R}^d$, $\mathcal{H} = \{\vec{x} \mid |\vec{x} - \vec{x}_0| \leq r\}$ spheres.

$VC(\mathcal{H}) < d+2$ because if we can put spheres around two disjoint sets, then those sets are also divided by a hyperplane, and $VC(\{\text{half-spaces}\}) = d+1$.

$VC(\mathcal{H}) \geq d+1$ by the same construction as half-spaces.

let $S = \{\vec{0}, \vec{e}_1, \ldots, \vec{e}_d\}$.

Given a subset $A \subseteq S$, choose the ball center $\vec{a}_0 = \sum_{a \in A} \vec{a}$.

Then $|\vec{a}_0 - a| = \sqrt{|A| - 1}$ $\forall\, a \in A$ and $a \neq 0$

$|\vec{a}_0 - a| = \sqrt{|A| + 1}$ $\forall\, a \notin A$ and $a \neq 0$.

$|\vec{a}_0| = \sqrt{|A|}$.

So we can choose a radius so that this ball contains exactly A.