

应用数学和信号处理相遇*

Stéphane Mallat著

陈刚 张亶 池哲儒 译

§ 1 超越Fourier变换

Fourier变换方法曾长时间地统治了信号处理的领域,留给新的、有挑战性的数学工具的只是不大的一点空间.七十年代以前,所考虑的大多数信号都是语音信号或其它声音信号,它们往往被认为是一个高斯过程.相应地,在几乎所有的过程中线性算法都被认为是最优的.因为是基于平稳性的假设,所以信号处理方法被严格地限制在一类卷积运算—Fourier变换中.

八十年代以来,随着图像处理研究的发展这种情况发生了彻底的变化.一般情况下,图像信号很难用高斯过程来进行刻画.并且图像中的一些突变结构例如边缘信息远比图像的平稳性重要.于是忽然间不可避免地涌现出了各种各样的非线性算法,从而也打开了信号处理通往现代数学的大门.除了传统的应用如信号传输、编码和信号恢复之外,信号处理也进入了信息分析的领域,语音理解和计算机视觉是其中的两大主要分支.信号处理和应用数学这种带有火花的相互碰撞也产生了许多新的数学问题.

对于信号处理和信息处理来说,为信号(函数)、处理方法(过程)和各种运算(算子)构造相应的稀疏表示式(sparse representations)是一个首要问题.一个稀疏表示式可以用特征化的较少的几个参数近似地表示信号,这些参数是从一组扩张的基或者一本冗长的“字典”中获得的.在稀疏表示的基础上,(关于信号处理的)复杂的非线性算法常常可以被简化和加速.稀疏表示同时也是处理很多数学问题的一个强有力的工具.在1990年的国际数学家大会上,Coifman和Meyer从调和分析的观点对此进行了分析;随后,在1994年的国际数学家大会上,Daubechies和Donoho又从数值分析和统计应用中紧支小波基的角度对稀疏表示式进行了解释.在共同研究**数据表示技术**的过程中,信号处理成为一种促进数学和工程界的重新交流与结合的推动力量.信号压缩、噪声滤波和随机模拟的各种应用导致我们从新的角度去研究逼近论、调和分析、算子理论、概率论和统计等数学分支并获得新的成果.

§ 2 稀疏表示式

稀疏表示式已经能够直接应用于数据压缩.但对于处理海量的数据信号而言,我们还需要降低分类和识别的复杂性.本节的叙述从逼近论开始,然后逐步把讨论的内容转移到信号压缩方面.

2.1 图像模型

Beyes认为,一个信号 $f(x)$ 可以被看作是某个过程 $F(x)$ 的一个样本函数(原文为realization),其误差的度量期望值与 F 的概率分布有关.事实上自然界的图像并不是一个高斯过程.至今还没有一个统计模型可以完全包容兼有边缘和各种纹理信息的复杂场景,图1(a)就是一个很典型的例子.这促使我们选择一个简单而实用的决定性模型:把一个信号看作 $L^2[0, 1]^d$ 中某个子集 S 中的一个函数 $f(x)$,但是我们没有任何关于这个子集 S 的概率分布的先验知识.接下来,对一个具体的过程,我们会试图极小化 S 中的信号对此具体过程的误差,这个方法称为minimax(极小极大)框架.在这样的模型中用 N 个采样点去离散一个信号 $f(x)$ 并不困难,因为在本质上就是将这个信号投影到一个 N 维的子空间上去.

很大一类图像,包括图1(a),都具有有界的全变差. $f(x)$ 在 $[0, 1]$ 的全变差被定义为 $f(x)$ 的变化率的振幅之和

$$\|f\|_{TV} = \int |f'(x)| dx < +\infty.$$

图像 f 在 $[0, 1]^2$ 上的全变差被定义为

$$\|f\|_{TV} = \iint |\nabla f(x)| dx < C.$$

这个范数有一个简单、基于水平集(level-sets)的几何解释:设

$$\Omega_t = \{(x, y) \in \mathbb{R}^2 : f(x, y) > t\}.$$

如果 $H^1(\partial\Omega_t)$ 是其边界 $\partial\Omega_t$ 的一维 Hausdorff 测度, 那么

$$\|f\|_{TV} = \int_{-\infty}^{+\infty} H^1(\partial\Omega_t) dt. \quad (1)$$

我们把有界变差与有界振幅合起来得到一个图像模型:

$$S_{BV} = \left\{ f : \|f\|_{TV} = \iint |\nabla f(x)| dx \leq C, \|f\|_{\infty} = \sup_{x \in [0,1]^2} |f(x)| \leq C \right\}. \quad (2)$$

这些图像具有典型的水平集和有限长度的“轮廓”. 这个模型虽然简单, 但已经足够用来体现信号表示的核心思想以及说明其困难所在. 更特别的一类图像, 比如具有同类纹理的图像, 可能用稀疏表示下的 Markov 随机域来表示更合适一些. 有关的内容在本文的第四节中介绍.

2.2 稀疏表示式和逼近论

设 $f \in L^2[0, 1]^d$ 在某个正交基 $\mathcal{B} = \{g_m\}_{m \in \mathbb{N}}$ 下的展开式为

$$f = \sum_{m=0}^{+\infty} \langle f, g_m \rangle g_m,$$

通过截断上述表达式可以获得 f 的一个稀疏表示式. 注意到研究某个函数在一组基下的稀疏表示也是逼近论中的一个中心课题. 在此我们仅简单介绍采用非线性逼近的动机, 更完整的叙述可以参见文献 [13].

设 V_M 是 \mathcal{B} 中 M 个向量 g_1, g_2, \dots, g_M 张成的子空间, 则对 f 用 M 个内积 $\langle f, g_m \rangle$ 的线性逼近就是 f 在空间 V_M 上的正交投影:

$$f_M = P_{V_M} f = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m.$$

这样, 在某个信号集合 S 上的极大逼近误差就是

$$\varepsilon_1(S, M) = \sup_{f \in S} \|f - f_M\|^2 = \sup_{f \in S} \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2.$$

如果随着 M 的增加 (此处原文为减少 (decrease), 疑为作者笔误. 译者注.) $\varepsilon_1(S, M)$ 衰减得非常快或者说随着 m 的增加 $|\langle f, g_m \rangle|$ 衰减得很快, 那么上式的逼近是十分有效的, 而这依赖于当 S 确定以后对 B 的选择. 例如, 一致正则函数就可以被 $L^2[0, 1]$ 中对应于 Fourier 基 $\{e^{i2\pi mx}\}_{m \in \mathbb{Z}}$ 的 M 个低频向量很好地逼近. 如果 S 被包含在周期为 1 的函数组成的 Sobolev 空间 $W^s[0, 1]$ 中的一个球中, 那么根据其 Fourier 系数在高频的衰减速度可以推出 $\varepsilon_1(S, M) = O(M^{-2s})$ (参见 [13]). 有界变差函数可能会不连续, 所以用 Fourier 基逼近的效果可能会不好. 利用 Kolmogorov 引入的 M -Width 的概念, 可以证明对于有界变差函数组成的一个球 S_{BV} , 在一组基 \mathcal{B} 下最快的误差衰减速度是 $\varepsilon_1(S_{BV}, M) \sim M^{-1}$ (参见 [13]).

为了改进以上的逼近效果, 可以建立一个更为自适应的表示: 把 f 投影到 M 个基向量上, 而这个 M 个基向量是依据 f 来选择的

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m. \quad (3)$$

因为 $\|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2$, 所以可以选择 M 个向量使得系数 $|\langle f, g_m \rangle|$ 具有最大振幅来得到最优逼近. 逼近的结果依赖于 $2M$ 个参数: 其中的 M 个是 I_M 中的向量, 而另外 M 个是内积 $\{\langle f, g_m \rangle\}_{m \in I_M}$. 我们

把这M个内积按递减排序(即当 $k \geq 1$ 时, $|c_k| \geq |c_{k+1}|$)并记为 $c_k = \langle f, g_{m_k} \rangle$. 这时非线性逼近的误差是:

$$\|f - f_M\|^2 = \sum_{k=M-1}^{+\infty} |c_k|^2, \quad \epsilon_M(S, M) = \sup_{f \in S} \|f - f_M\|^2.$$

误差值与 $|c_k|$ 的衰减速度有关. 对于上述的基 \mathcal{B} , 我们定义一个半径为C的 wI^p 球

$$S_{wI^p} = \{f: |c_k| = |\langle f, g_{m_k} \rangle| \leq Ck^{-1/p}\}. \quad (4)$$

我们很容易验证: 存在某个 $p < 2$ 和 $C > 0$ 使得 $S \subset S_{wI^p}$ 成立的充要条件是 $\epsilon_n(S, M) = O(M^{-\frac{2}{p}})$. 对于非线性逼近来说, 主要的困难是寻找最小的 p 和相应的基 \mathcal{B} 满足 $S \subset S_{wI^p}$. 这样得到的一组基被称为S的一个最优非线性逼近基. 无条件基就是最优基的一个例子.

Banach子空间 $B \subset L^2[0, 1]^d$ 中的一组正交基 \mathcal{B} 被称为无条件基, 如果存在正常数A, 对于任意的符号序列 $s_m \in \{-1, 1\}$ 和 $f \in B$, 有

$$\left\| \sum_{m=0}^{+\infty} s_m \langle f, g_m \rangle g_m \right\|_B \leq A \left\| \sum_{m=0}^{+\infty} \langle f, g_m \rangle g_m \right\|_B.$$

注意到 $\|f\|_B < +\infty$ 可以用振幅 $|\langle f, g_m \rangle|$, 以及相应的、经排序后系数的衰减条件来刻画. 从而人们可以证明下面的事实^[13]: 如果 \mathcal{B} 是空间B中的无条件基, 那么对B中的一个球 $S = \{f: \|f\|_B \leq C\}$ 的非线性逼近而言, \mathcal{B} 是最优基.

2.3 自适应的小波网格

小波基在数学和信号处理中都有广泛的应用, 这是因为小波具有极强的能力去构造一大类函数的稀疏表示式. $L^2(\mathbb{R})$ 中的第一个正交小波基是由Strömberg和Meyer^[25]首先提出的. 利用多分辨率分析框架, 我们几乎可以构造所有生成 $L^2(\mathbb{R})$ 的小波基^[19]. 同样, 在上述框架下, 文献^[22]给出了一个快速的离散算法: 计算N个小波系数所需要的计算量仅为 $O(N)$. Daubechies^[9]发现了具有紧支集的小波, 这个结果已经被推广到了 $L^2[0, 1]^d$. 在1994年的国际数学家大会上^[10], 她报告了有关的主要结果. 在此我们仅作一个简要的叙述.

$L^2[0, 1]$ 中的正交小波基是一族函数

$$\mathcal{B} = (\{\phi_{l,n}\}_{0 \leq n < 2^l} \cup \{\psi_{j,n}\}_{j \geq 1, 0 \leq n < 2^j}).$$

在 2^1 级分辨率上, 尺度函数 $\{\phi_{1,n}\}_{0 \leq n < 2^1}$ 可以生成一个空间 V_1 , 对所有的 $q \geq 0$, V_1 包含了所有的 q 次多项式. 而在更高的 2^j ($2^j > 2^1$)级分辨率下, 小波 $\psi_{j,n}$ 正交于所有的 q 次多项式. 可以通过伸缩和平移“母小波” ψ 来生成支集在 $(0, 1)$ 上的小波族 $\psi_{j,n}$

$$\psi_{j,n}(t) = \sqrt{2^j} \psi(2^j t - n).$$

并且可以通过某些方法, 使边界小波²(靠边界的小波被称为边界小波(Boundary wavelets), 译者注.)的支集落入 $[0, 1]$.

一个函数(信号) f 用M层小波($M=2^J > 2^j$)和尺度函数去线性逼近的过程是保留所有 $2^j < 2^1$ 级分辨率上的小波系数. 即

$$f_M = \sum_{n=0}^{2^1} \langle f, \phi_{1,n} \rangle + \sum_{j=1}^{J-1} \sum_{n=0}^{2^j} \langle f, \psi_{j,n} \rangle \psi_{j,n}.$$

上式中第一部分的和式表示 f 在 2^1 的分辨率上的一个粗糙逼近, 而每一个部分和 $\sum_{n=0}^{2^j} \langle f, \psi_{j,n} \rangle \psi_{j,n}$ 表示从分辨率 2^j 到 2^{j+1} 的“细节”, 可以用它们逐步改善逼近的效果. 如果 f 是连续的, 在 2^J 级分辨率上的线

性逼近实际上等价于对采样点 $\{f(2^{-j}n)\}_{0 \leq n < 2^j}$ 的均匀网格插值逼近. 类似于线性Fourier逼近, 只有当 f 是一致正则函数时均匀网格逼近的效果才是好的. 对于包含奇异信息的函数如有界变差函数来说这种逼近的效果是糟糕的.

非线性小波逼近保留了 M 个振幅(绝对值)最大的小波系数, 而 $|\langle f, \psi_{j,n} \rangle|$ 的振幅大小取决于 f 的局部正则性. 假设“母小波” ψ 具有 C^{q+1} 阶光滑性, 并且正交于 q 次多项式, 可以证明 [25], f 在区间 $[a, b]$ 上一致地具有 Lipschitz 指数 $\alpha < q+1$ 的充分必要条件是: (模边界问题) 存在 $A > 0$, 对所有支集包含在 $[a, b]$ 中的非边界小波 $\psi_{j,n}$, 都有

$$|\langle f, \psi_{j,n} \rangle| \leq A 2^{-(\alpha+1/2)j}.$$

在 Lipschitz 系数 α 较大的区域, 当分辨率 2^j 增加时 $|\langle f, \psi_{j,n} \rangle|$ 衰减得非常快; 当 $0 \leq \alpha < 1$ 时, 振幅大的小波系数总出现在奇异点的一个邻域内. 作非线性逼近的时候, 在奇异点附近更多的小波系数被保留下来. 所以非线性小波逼近就等价于一个自适应的网格逼近, 网格的分辨率在奇异点的邻域内应当加细.

小波基的紧性在泛函分析中的意义来源于这样一个事实: 它是一大类光滑空间 (Besov 空间) 的无条件基, 因此对这些空间中的球所作的非线性逼近是最优的. 虽然有界变差函数空间 BV 不存在无条件基, 但它可以被嵌入到两个 Besov 空间中. 这样我们可以证明用小波基去逼近有界变差函数空间中的球 S_{BV} 是最优的. 当 $p=2/3$ 时, 一个球 S_{BV} 被包含在 wI^p 球 (参见 (4) 式) 中, 但当 $p < 2/3$ 时, 这个命题不成立. 因此 $\epsilon_n(S_{BV}, M) = O(M^{1-2/p}) = O(M^{-2})$. 当 M 增加的时候, 用小波基作非线性逼近时 $\epsilon_n(S_{BV}, M)$ 的渐近衰减比任何用 M 个参数的线性逼近都快, 因为线性逼近的渐近衰减的速度至多为 M^{-1} .

在二维的情况下, 可以用三个“母小波”函数 ψ^k ($1 \leq k \leq 3$) 来构造小波基, 它们经过伸缩和平移生成

$$\psi_{j,n}^k(x_1, x_2) = \psi_{j,n}^k(x) = 2^j \psi^k(2^j x_1 - n_1, 2^j x_2 - n_2).$$

在边界附近作适当的修改后, 可使它们的支集落在 $[0, 1]^2$ 中. 那么小波 $\psi_{j,n}^k$ 的方形支集的边长与 2^{-j} 成比例, 其中心靠近 $2^{-j}n = (2^{-j}n_1, 2^{-j}n_2)$. $L^2[0, 1]^2$ 中的一个正交小波基可以通过不断增加低分辨率的尺度函数而得到

$$B = (\{\varphi_{1,n}\}_{2^{-1}n \in [0, 1]^2} \cup \{\psi_{j,n}^k\}_{j \geq 1, 2^{-j}n \in [0, 1]^2, 1 \leq k \leq 3}). \quad (5)$$

一幅离散的图像由 N^2 个点 (象素) 组成, 在图 1(a) 中, $N=512$. (5) 式中的小波基可以被离散化为 N^2 象素的图像的正交基. 图 1(a) 经小波变换的结果如图 1(b) 所示. 图 1(b) 的每个子图像中, 每给定一组 j 和 k 就对应一个值 $|\langle f, \psi_{j,n}^k \rangle|_{2^{-j}n \in [0, 1]^2}$. 在每个子图像中的小波系数 (象素) 的数目是 2^{2j} , 白和黑 (灰度的程度) 的点分别表示了该点的值是接近于零还是取较大的 (振幅) 值. 这些子图像中每三个被分为一组, 分别对应了 $1 \leq k \leq 3$ 的取值, k 值不同的小波对应了图像沿不同方向的变化情况. 可以看到图像中大部分点是白的, 说明对应的大部分小波系数是接近于 0 的. 只有较少的小波系数值比较大, 它们反映的是图像灰度值变化比较大的区域, 如图像的“边缘”和“纹理”区域.

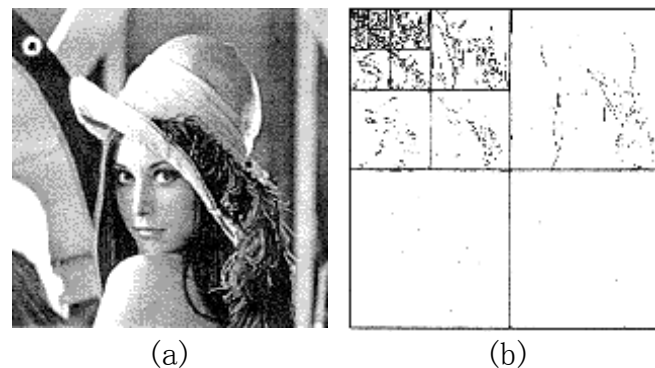


图1 (a)原图像 f . (b) 在正交小波基下系数 $|\langle f, \psi_{j,n}^k \rangle|$ 的量化表示.

每一个子图对应于一个不同的分辨率 2^j 和不同的方向 k (参见正文).

用 M 层小波($M=2^{2J}$)作线性逼近的方法是保留所有 $2^j < 2^1$ 级分辨率上的小波系数. 对于不连续的信号, 均匀的网格逼近是失效的. 对于有界变差界图像空间(2)中的一个球, 可以证明 $\varepsilon_1(S_{BV}, M) = A > 0$, 极大逼近误差并没有随着 M 的增加而趋向于0.

非线性的逼近比线性逼近更为有效是因为它保留了奇异点附近的小波系数, 并且(1)式说明“图像边缘”的长度是有限的. 更精确地说, 可以证明 S_{BV} 被包含在一个 wI^1 球中, 并且逼近误差 $\varepsilon_n(S_{BV}, M) = O(M^{-1})$. 对图像而言, 小波自适应网格逼近比均匀逼近的效果好出许多, 没有其它一种正交基的逼近阶可以超过正交小波基.

2.4 信号压缩

在有带宽限制的信道(例如Internet)中对海量数据的信号进行经济地存储和快速的传输是信号压缩的主要应用领域. 提高编码(压缩)效率, 也就是用尽可能少的位(bit)数表示一个信号需要给出信号的稀疏表示式. 为了满足在语音和图像压缩中的需要, 信号处理领域的工程师们没有等数学中非线性逼近理论的成熟就开始了小波基的应用. 1986年有了第一个小波图像编码器^[34], 那时在数学领域中对于小波正交基的研究还没有真正开始. 所谓的快速正交小波变换实质上采用的是一个被称为“滤波器组”(Filter Bank)的算法. 这个算法最初被用在复路信号(利用一个信道传输多路信号)的处理上^[7]. 文献^[33]首次将Filter Bank的理论引入到信号处理中, 但在后来的文献^[19]中才真正将正交小波基与之相联系. 虽然数学家们姗姗来迟, 但在分析图像编码器的工作机理方面确实需要最新的逼近论成果, 这样一来也为该领域的潜在发展开辟了众多新的方向.

S 中的信号在 N 级分辨率上被离散和逼近, 则意味着该离散信号属于一个 N 维空间. 信号 f 在正交基 $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ 下的变换编码定义为

$$f = \sum_{m=0}^{N-1} \langle f, g_m \rangle g_m,$$

和每个系数 $\langle f, g_m \rangle$ 的量化值来近似, 量化值一般用尽可能少的位数进行编码. 一个步长为 Δ 的均匀量化器用 $Q(x) = k\Delta$ 逼近 $x \in \mathbb{R}$, 其中 $k \in \mathbb{Z}$, $|x - Q(x)| \leq \Delta/2$. 量化的结果是:

$$\tilde{f} = \sum_{m=0}^{N-1} Q(\langle f, g_m \rangle) g_m.$$

我们需要解决的问题是如何分配至多 R 位的存储量对 \tilde{f} 进行编码并获得最小的失真 $d(S, R) = \sup_{f \in S} \|f - \tilde{f}\|^2$.

变换编码的失真率首先与非线性逼近有关. 假设绝对值超过 $\Delta/2$ 的系数数目是 M 并且 f_M 是用最大的 M 个系数所得到的 f 的非线性逼近

$$f_M = \sum_{|\langle f, g_m \rangle| > \Delta/2} \langle f, g_m \rangle g_m.$$

因为当 $|x| < \Delta/2$, $Q(x) = 0$, 并且 $|x - Q(x)| \leq \Delta/2$, 所以失真

$$d(f, R) = \|f - \tilde{f}\|^2 \leq \|f - f_M\|^2 + M(\Delta^2/4). \quad (6)$$

这启发我们用非线性逼近方法考虑失真问题. 假设 S 在一个半径为 C 的 wI^p 球 S_{wI^p} (4)中, 这里 $p < 2$. 记 $M_0 = C^p (\Delta/2)^{-p}$, 因为 $|c_k| = |\langle f, g_{m_k} \rangle| \leq Ck^{-1/p}$, 必要的话, 让 $M \leq M_0$, 则可验证

$$d(S, R) = \sup_{f \in S} \|f - \tilde{f}\|^2 \leq \sup_{f \in S} \|f - f_M\|^2 + M_0 \Delta^2/4 = O(M_0^{-2/p+1}). \quad (7)$$

所产生的总量化失真可用非线性逼近的误差来控制.

为了优化变换编码, 我们必须对表示信号 f 的 N 个值 $\{Q(\langle f, g_m \rangle)\}_{0 \leq m < N}$ 进行编码, 并使得所需要

的最大存储量(比特位)尽量小. 在高压缩率的情况下, 即 $N \cdot M_0 \geq M$, 此时, 占比例为 $\frac{N-M}{N}$ 的大部分系数数值都被量化为0. 熵编码正是利用了这个特性, 它采用较少的位数编码出现概率较高的系数. 如果已知 $S \subset S_{w1p}$, 那么可以构造最大比特数为 $R \sim M_0 \log_2 \frac{M_0}{N}$ 的算术编码方法^[22]. 并且从(7)式可以导出, 对任意的 $q > p$, 其量化失真为 $d(S, R) = O(R^{1-2/q})$.

奇妙的是, 当采用的是使S的非线性逼近最优的基B时, $d(S, R)$ 的衰减率也同时达到最快, 因为此时选择的是满足 $S \subset S_{w1p}$ 的最小的指数p. 特别是, 对于有界变差图像来说, 小波基是最优的并且极小指数为 $p=1$. 图2(a)和图2(b)是对 \tilde{f} 进行量化(压缩)后的解码图像, 而 \tilde{f} 中通过对图1(b)中的小波系数进行量化后得到的. 对应编码率R/N分别为0.25bits/像素和0.125bits/像素, 采用的是对零系数的最优编码器^[30]. 注意到图1(a)是用8bits/像素表示的, 所以它们对应的压缩比分别为32倍和64倍. 对0.25bits/像素的解码图像, 很难察觉视觉上的失真, 但对于0.125bits/像素的解码图像就比较明显.

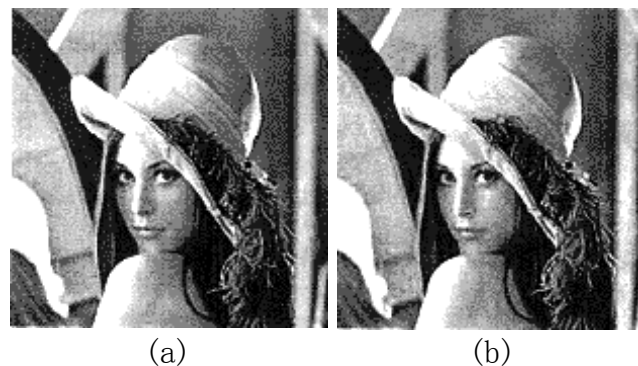


图2 (a)对图1中的原图用量化小波系数的方法得到的编码图像, 码率为0.25/像素. (b)码率为0.125/像素的编码图像.

我们需要强调, 小波基的选择完全依赖于S中信号的特性. 比如对于声音信号, 选择各种不同的基都是为了有效地逼近各种复杂的振荡波形. 图3显示的是单词“greasy”的录音信号. 最近的关于CD的语音压缩标准例如杜比AC系统在计算时采用了与局部余弦函数类似的基. 这个基通过一个称为窗口的偶函数 $\omega(t)$ 进行构造, 它具有支集 $[-21, 21]$, 并且 $\omega(t)$ 的平移均匀地覆盖了整个实轴:

$$\sum_{p=-\infty}^{+\infty} |\omega(t - pl)|^2 = 1.$$

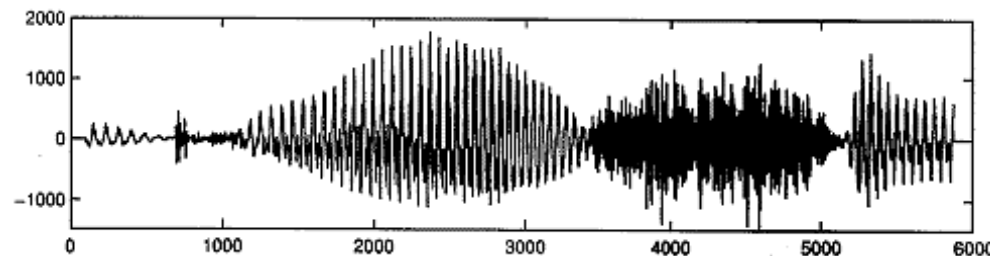


图3 录制并以16kHz采样的单词“greasy”的语音信号

Malvar^[23], Coifman和Meyer^[5]证明了, 如果给 $\omega(t)$ 附加上一些对称性后再乘以余弦函数, 那么就可以形成 $L^2(\mathbb{R})$ 的一个正交基:

$$\{g_{p,k}(t) = \frac{1}{\sqrt{l}} \omega(t - pl) \cos(\pi k(l^{-1}t - p))\}_{k \in \mathbb{N}, p \in \mathbb{Z}}. \quad (8)$$

与处理图像时的情况一样, 在对语音信号进行编码过程中选用这组基是因为可以用很少的几个局部余弦向量来逼近声音信号. 但是在很多情况下语音编码的失真不是用 L^2 范数来衡量的, 工程师们采用复杂的掩盖技术可以达到人的主观听觉所无法觉察的量化失真. 当然这已经超出了数学的范围.

2.5 自适应的几何特性和其他特性

小波基用来表示一般的有界变差图像可以达到最优, 但为了获得更好的逼近效果则需要利用大多数图像所具有的几何正则性. 全变差公式(1)告诉我们: 有界变差图像的水平集基本上都具有有限长度.

但这并没有对水平集附加任何正则性条件. 在图1(a)中, 大部分“轮廓”是由图像平面上的分段正则几何曲线组成的, 这些曲线在大多数位置都具有比较小的曲率. 如何利用图像的这种正则性是图像处理的一个基本问题. 已有研究者采用非线性偏微分方程的方法去改变图像水平集的曲率 [1, 28, 31]. 这种把数学方法运用于图像处理的新的重要的尝试, 导致了诸如去除图像噪声和图像分割等方面的一些非常有意义的应用. 当然, 关于这方面的进一步叙述会偏离本文的主题, 我们的讨论目标依然是稀疏表示式.

为了说明几何正则性的重要性, 让我们先观察一个简单的“图像” $f=1_{\Omega}$, 它是集合 Ω 的特征函数. Ω 的边界 $\partial\Omega$ 是一条可微的、有限长度和曲率有界的曲线. 如果 $\psi_{j,n}^k$ 的方形支集和 $\partial\Omega$ 不相交, 那么 $\langle f, \psi_{j,n}^k \rangle = 0$. 小波 $\psi_{j,n}^k$ 在步长为 2^{-j} 的方形网格上平移, 并且其方形支集的边长与 2^j 成比例, 如图4(a)所示. 在 2^j 级分辨率上, 有 $O(2^j)$ 个小波的支集和 $\partial\Omega$ 相交. 选择其中振幅最大的 M 个小波系数去做非线性逼近, 而没有被选中的小波产生了误差 $\|f-f_M\|^2 \sim M^{-1}$, 这样的做法对于一般有界变差图像也类似.

更好的方法是选用 $[0, 1]^2$ 中 M 个自适应的三角片来进行分段线性逼近 [16]. 因为 $\partial\Omega$ 的曲率是有界的, 所以必定可以被 $M/2$ 个三角片所覆盖, 这些三角片都很窄, 其沿着 $\partial\Omega$ 的法线方向的宽度正比于 M^{-2} , 但沿着 $\partial\Omega$ 的切线方向可被拉长. Ω 是外部和内部可被 $M/2$ 个较大的三角片覆盖, 如图4(b)所示. 在每个三角片上都是线性函数的函数 f_M 可按误差 $\|f-f_M\| = O(M^{-2})$ 来逼近函数 $f=1_{\Omega}$. 逼近误差集中在边缘上, 三角片的窄边导致了逼近误差比采用小波网格的逼近误差更小. 因为三角片对于 $\partial\Omega$ 的几何形状是自适应的, 所以这个误差会随着分辨率提高而下降.

建立几何特性和自适应逼近之间的联系是图像处理的一个基本问题. 人的视觉系统特别关注几何特征例如“转角”以及边界的规则性 [24] [26]. 图5中的Kanizsa错觉就说明了这样一个事实. 我们意识到了一个三角形“边界”的存在虽然在图像的中央区域并没有灰度的变化. 这些例子说明人类在感知一幅图像时带有与生俱来的几何限制. 人类的视觉皮层中简单细胞对所观察图像的处理相当于用一类函数对图像进行分解, 这类函数与小波非常接近, 但是细胞的处理过程中冗余度更少并且更灵活一些. 以上的叙述说明了我们的大脑可以建立跨越泛函分析和几何之间的桥梁.

自适应地利用图像的几何特性是一个更广泛的自适应逼近问题. 一组基在我们所使用的函数空间中是完备的, 但是对表达复杂信号中所包含的各种结构而言这组基往往显得太小了. 从一个大的“基库 (原文为dictionary, 此处根据意思译为“基库”.)” $D=\{g_r\}_{r \in \Gamma}$ 中选择 M 个向量就会获得更加精确的逼近, 而这个“基库”可能包含有无穷多组基. 这就类似于人在掌握了更大的词汇量后就可以造出更简明、更贴近的句子. 对于模式识别来说, 寻求具有平移和仿射不变性的表示也是十分重要的. 这就要求给“基库”加上更多的条件 [20]. 图像的一个“基库”可以用支集长度和方向参数化的一类小波来构造. 就象图4(b)中可变长的三角片, 所选择的小波可以适应于图像的水平集. 声音信号也可由局部余弦向量组成的“基库”有效地逼近 (参见(8)式), 但是其中窗口的长度 l 可以自适应于波形的变化, 因为波形是由打击声、音乐或者其他短暂的声源所产生的, 其持续时间各有不同.

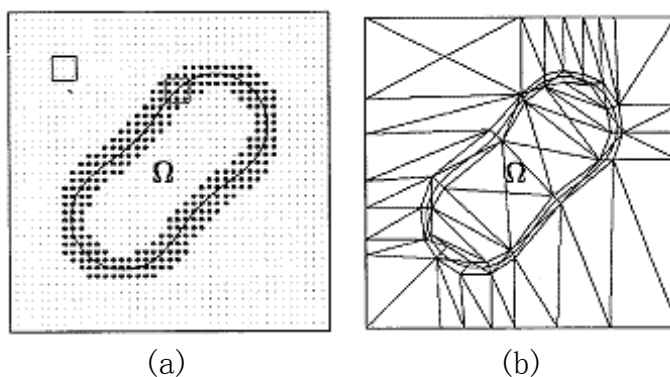


图4 (a) 通过在区间 2^{-j} 上正方形网格平移形成的小波 $\psi_{j,n}$, 其正方形支集的边长正比于 2^{-j} . 对 $f=1_{\Omega}$, 图中较暗的点表示的使得 $\langle f, \psi_{j,n} \rangle \neq 0$ 的小波点. (b) 对 $f=1_{\Omega}$, 当沿着 f 在边界上的连续点选取拉长的窄边三角形时, 这样的分段线性逼近是最优的.



图5 视觉可以感知到的其实并不存在的完全由白色点组成的直边和曲边三角形图像。

一个自适应的表达式就是从“基库”D中选择M个向量 $\{g_{\gamma_k}\}_{1 \leq k \leq M}$ 后用一个部分和去逼近f

$$f_M = \sum_{k=1}^M \alpha_k g_{\gamma_k}.$$

如果失去正交性，寻找M个向量使得 $\|f - f_M\|$ 最小的过程可能会导致组合爆炸. 贪婪算法(Greed Pursuit Algorithm)避免了组合爆炸^[20]，它采用的方法是从“基库”中逐个选取向量 g_{γ_k} ，但是它们的逼近效果远不是最优的^[3, 13]. 在采用由具有树结构的正交基组构造“基库”过程中，Coifman和Wickerhauser引入了动态规划算法^[6]，使得所选取的M个向量是“合理的”，但还不是最优的. 至今还没有一种逼近理论可以分析这样的非线性逼近过程并设法作出改进.

我们最后必须提及的是，扩大“基库”的容量需要代价. 在一个较大的“基库”中，需要更多的参数去刻画被选择向量的指标 γ_k . 对于确定的逼近误差而言，寻找一个太大的“基库”需要增加大量参数去表征信号的逼近函数 f_M . 因此，找一个大小适中的“基库”成了另外一个公开的难题.

§ 3 用阈值法去除噪声

在测量或传输的过程中，噪声会掺杂到信号中. 因此去除噪声是一个十分重要的问题，而稀疏表示在去除噪声方面扮演着重要的作用. 当一个信号被变换到一组基上后，这个信号被分解为若干个“大振幅”的值再加上一个小的余项. 通过设置一个合适的阈值可以把小(振幅)系数置为0，这样可以比较容易地去除噪声. 六十年代起，类似的想法已经被使用在电视信号的噪声去除上. 但直到最近Donoho和Johnstone^[14]才从数学上证明了这样的阈值估计器在稀疏表示中几乎是最优的，从而开辟了信号处理中的许多新应用.

假定用N个系数去离散地逼近定义在 $[0, 1]^d$ 上的函数 $f(x)$ ，我们把它记为 $f[n]$ ， $0 \leq n < N$. 设被测量的含有噪音的数据是

$$D[n] = f[n] + W[n], \quad (9)$$

其中噪声 $W[n]$ 被认为是一个独立的Guass随机变量，因此也可以把它看成一个白噪声. 图6(a)是一个例子. f 的一个估计器(estimator)F就是把一个算子L作用在数据上， $F = LD$. 估计的偏差(risk)是 $r(L, f) = E\{\|f - LD\|^2\}$. 我们要把整个信号集合S偏差的最大值极小化 $r(L, S) = \sup_{f \in S} r(L, f)$. 也就是要找一个算子L使得其作用的偏差达到最小 $r_0(S) = \inf_{L} r(L, S)$. 这样的问题在数理统计领域已经有相当数量的文献进行研究.

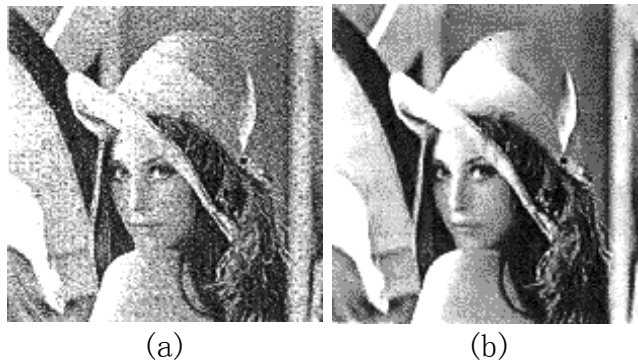


图6 (a)被叠加Guass白噪声污染的图像. (b)在小波基下用阈值估计法去除噪声后的结果.

一个新的方法就是把估计式和表达式分开. 第一步是在一组正交基 $\mathcal{B} = \{g_m\}_{0 \leq m \leq N}$ 下通过分解 $D = f + W$ 得到一个合适的表达式:

$$\langle D, g_m \rangle = \langle f, g_m \rangle + \langle W, g_m \rangle.$$

一个阈值估计器的简单定义如下:

$$F = L_t D = \sum_{n=0}^{N-1} \theta_T(\langle D, g_n \rangle) g_n, \quad (10)$$

其中 $\theta_T(x) = x \mathbf{1}_{|x| > T}$. 于是把低于阈值 T 的系数置为 0 而保留其他的系数. 选择阈值 T 使得它在大概率的意义上恰好大于 $\max_{0 \leq m < N} |\langle W, g_m \rangle|$, 因此如果 $\langle f, g_m \rangle \approx 0$, 那么 $\theta_T(\langle D, g_m \rangle) = 0$.

因为 W 是一个方差为 σ^2 的 Gauss 白噪声, 所以对于任何基 B , 噪声系数 $|\langle W, g_m \rangle|$ 也是方差为 σ^2 的独立 Gauss 随机变量. 假设 M 是满足 $|\langle f, g_m \rangle| \geq \sigma$ 的系数个数, f_M 是由这 M 个向量按 (3) 得到的非线性逼近表达式. Donoho 和 Johnstone 证明了如果 $T = \sigma \sqrt{2 \ln N}$, 那么 $r(L_t, f) \leq (2 \ln N + 1) (\|f - f_M\|^2 + (M+1)\sigma^2)$. 上式右边的上界与变换编码的失真估计 (6) 类似. 我们看到, 如果能选择一组基使得超过 σ 的大振幅系数比较小, 同时逼近误差 $\|f - f_M\|$ 也比较小的话, 估计偏差就会很小. 我们需要面对的另一个问题是构造一个既稀疏又精确的表达式. 图 6 (b) 是对加噪声的图 6 (a) 的小波系数用阈值法进行估计后的还原像.

也可以估计在分辨率 $N \rightarrow +\infty$ 时阈值估计器的渐近性质. 给定集合 S_0 中的一个信号 $f(x)$, 寻找一组对于非线性逼近来说是最优的正交基. 假设 S_0 是空间 B 中的一个球, 我们选择 \mathcal{B}_0 是 B 的无条件基. 离散信号集合 S 可以通过在 N 维空间上进行投影得到. 通过相同的投影, 这些信号在由 \mathcal{B}_0 导出的基 \mathcal{B} 下被分解. 当 N 增加时, 阈值估计在下面的意义上是几乎最优的

$$r(L_t, S) \leq O(\log N) r_0(S).$$

这个结果可以被应用到 Besov 空间中的信号的离散化, 以及此后在离散小波变换的分解. 这样的方法对于有界变差信号集 S_{BV} 也是有效的, 因为 BV 可以被嵌入两个到非常靠近的 Besov 空间中. 在这种情况下, 当 N 增加时, 阈值估计的偏差渐近衰减的速度比其他的任何线性估计都要快.

阈值估计的效果关键取决于表达式的逼近效果. 为了利用复杂信号中的结构性信息, 如有些图像所具有的几何正则性, 阈值估计必须利用更为自适应的表达式进行计算, 参见 2.5 节. 但是, 这种高度自适应估计器的极大极小化问题仍需进一步理解.

§ 4 稀疏交互处理

在包括语音识别和纹理描述的许多关于分类的问题上, 所观察到的信号往往被描述成一个我们需要根据信号的特性来进行刻划的某个过程的样本点. 因为所描述的过程通常不是 Gauss 过程或者是非平稳过程, 所以通过单一的模型所提供的少量数据来进行刻画是非常困难的. 我们需要同时用几个参数在合适的表达式中来刻划这些过程, 以便这些过程能够被更好地用于估计和分类. 在考察非平稳的 Gauss 过程后, 我们将讨论更为一般的 Markov 随机域模型.

4.1 非平稳的 Gauss 过程

Gauss 过程为包括语音信号在内的一大类信号提供了合理的模型. 设 $t \in \mathbb{R}$, 一个零均值的 Gauss 过程 $X(t)$ 可被它的协方差 $\kappa(t, s) = E\{X(t)X(s)\}$ 唯一地确定, 其中协方差是协方差算子 K 的核函数 (kernel),

$$Kf(t) = \int_{-\infty}^{+\infty} \kappa(t, s) f(s) ds. \quad (12)$$

为了从少数几个样本函数中估计协方差, 必须减少确定核函数的参数个数. 这可以通过寻找一组正交基 $\mathcal{B} = \{g_m\}_{m \in \mathbb{Z}}$ 来实现, 这组基下所对应的矩阵在非对角线上的元素是速降的, 矩阵的元素是

$$\langle Kg_n, g_n \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \kappa(t, s) g_n(s) g_n(t) ds dt. \quad (13)$$

实际上, 上述矩阵的元素就是核函数 $\kappa(t, s)$ 按 $L^2(\mathbb{R}^2)$ 中一组可分离的正交基 $\{g_n(t)g_m(s)\}_{(n, m) \in \mathbb{Z}^2}$ 分解所得到的系数. 找一个稀疏矩阵的表示因此也就等价于在一组可分基下用较少的非零系数逼近 $\kappa(t, s)$. 如果矩阵的非对角线元素衰减得足够快, K 就能在极大模或者 Hilbert-Schmidt 模意义下被 \mathcal{B} 中一个窄带矩阵 \mathbf{K} 很好的逼近, 其中 \mathcal{B} 是用来逼近 X 的 Gauss 过程 \mathbf{X} 的协方差矩阵 [21]. 因为 \mathbf{K} 具有带状矩阵表

达式, 所以对每个 $m \in \mathcal{N}$, 存在一个由有限个整数组成的一个邻域 \mathcal{N}_m 满足如果 $n \notin \mathcal{N}_m$, 那么

$$\langle \tilde{K} g_m, g_n \rangle = E\{\langle \tilde{X}, g_m \rangle \langle \tilde{X}, g_n \rangle\} = 0.$$

因为 $\langle \tilde{X}, g_m \rangle$ 和 $\langle \tilde{X}, g_n \rangle$ 是同分布的Gauss随机变量, 而它们互不相关, 所以是独立的. 作为X的模型, \tilde{X} 在 \mathcal{D} 中的表达式系数仅取决于一个很小的领域, 这正是Markov随机域的一个例子.

把协方差算子K看作是一个拟微分算子去寻找一组基, 使得在这组基下的矩阵是非对角线速降的, 这是一个很有用的方法 [25]. 令 $\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(s)e^{-i\omega s} ds$ 是f的Fourier变换. 那么算子K的符号是

$$\beta(t, \omega) = \text{p. v.} \int_{-\infty}^{+\infty} k(t, t-s)e^{-i\omega s} ds.$$

对(12)使用Parseval公式有

$$Kf(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \beta(t, \omega) \hat{f}(\omega) e^{i\omega t} d\omega.$$

举例来说, 如果 $\beta(t, \omega) = \sum_{p=0}^P a_p(t) (i\omega)^p$, 那么 $K = \sum_{p=0}^P a_p(t) \left(\frac{d}{dt} \right)^p$ 是一个关于时变系数的微分算子. 如果 $\kappa(t, s) = \kappa(t-s)$, 那么随机过程X是平稳的, 并且 $\beta(t, \omega) = \beta(\omega)$ 是K的谱. 因此Fourier变换是描述平稳Gauss过程的一个理想工具. 对于非平稳的随机过程, 我们需要建立X(t)和 $\beta(t, \omega)$ 之间的联系, 并且得到一组基使得K可被一个窄带矩阵逼近.

局部的平稳过程X(t)出现在许多物理系统中. 这些系统的某些力学机制导致该过程在时间或者空间中出现随机扰动的变化非常缓慢. 在一个很短的时间段1上, 这样的过程可以被一个平稳的随机过程来逼近. 许多语音或声音信号就属于这种情况. 在充分短的时间段内, 喉咙就象一个稳定的共振器, 由平稳的音源所激发. 一类简单的局部的平稳过程可以通过附加这样一个条件而得到: 存在 $A > 0$, 使得对于所有的k, $j \geq 0$

$$|\partial_t^k \partial_\omega^j \beta(t, \omega)| \leq A t^{j-k}.$$

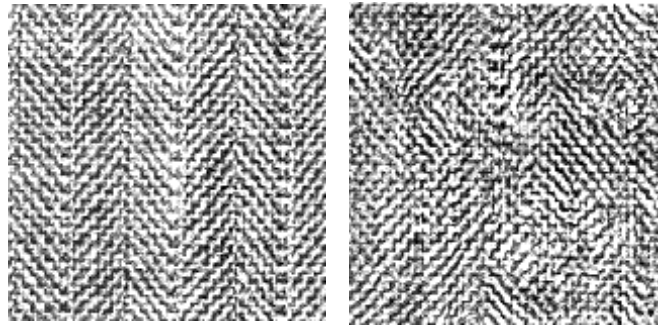
我们导出局部余弦基(8)的存在性, 它使得算子K可以非常有效地用一个窄带矩阵来逼近(参见 [13]). 窗口尺度1可以自适应于稳定性区间. 当稳定性区间的长度1(t)随时间变化非常剧烈的时候, 比如声音信号, 那么相应的协方差算子会具有较强的复杂性, 并且这个算子不属于某一类拟微分算子. 依照1(t)的规律, 自适应的局部余弦基依然也为这类算子提供了稀疏表示 [21].

多重分形为分析具有自相似性的信号提供了很有用的模型. 我们可以举许多这样的例子, 如经济领域中的道琼斯工业平均指数、心电信号数据、银河系中某辐射源的电磁脉冲、某些图像的纹理、交通流量的变化..., 等等. Hurst指数为H的分形布朗运动X(t)是分形Gauss过程的一个典型例子, 它的增量是稳定的, 并且在下列意义下是自相似的: 对于所有的 $s > 0$, $s^{-H} X(st)$ 与X(t)具有相同的概率分布. X的协方差K的符号是 $\beta(t, \omega) = \lambda |\omega|^{-2H-1}$. 它对应了第一代的Calderon-Zygmund算子, 在小波基下所对应的矩阵是非对角线速降的. 在信号处理领域, 分形布朗运动往往可被某个随机过程 \tilde{X} 所逼近, 其中 \tilde{X} 所对应的协方差在小波基下是对角的, 这导致了一个快速的综合算法 [27]. 通过对 $\partial_t^k \partial_\omega^j \beta(t, \omega)$ 加上一般性条件可以保证K在小波基下是非对角线速降的. 多重分形布朗运动提供了Hurst指数随时间变化的例子: $\beta(t, \omega) = \beta_0(t) |\omega|^{-2H(t)-1}$. 在小波基下也可对 $\beta_0(t)$ 和H(t)作精确估计.

当随机过程是一致局部平稳的或者多重分形的时候, 用来压缩协方差矩阵的基是事先知道的. 但对于一些更复杂的非平稳随机过程, 这组基必须依据先验信息进行估算. 如同2.5节中描述的那样, 这也是一个自适应逼近的问题, 尽管逼近算子与函数并非是一致的. 最优基自适应搜索算法被用来逼近协方差算子的技术目前还处于萌芽阶段, 进一步的工作需要我们对统计数据的估计有一个充分的了解.

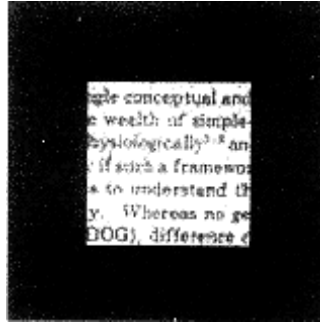
4.2 稀疏表示中的Markov随机域理论

纹理的特征化描述和综合处理是底层可视化的一个最具有挑战性的问题. 同样性质的纹理例如材质、地毯或者大理石的图像可以被看成是一个平稳过程, 但却不是Gauss过程. 图7给出了两个这样的例子. 我们需要用极少的几个参数来建立这些过程的模型, 并希望能在一个简单的过程中识别它们. 人类的视觉系统可以灵活地完成这个功能. 这个问题的重要性已经超出了单纯的纹理描述. 实际上, 有必要建立一个一般性的框架去模拟非Gauss过程, 以便对诸如金融时间序列和湍流的速率变化等各种不同类型信号进行分析.

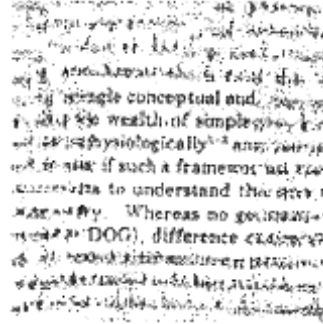


(a)

(b)



(c)



(d)

图7 (a) 观察到的均匀纹理. (b) 对(a)利用小波Markov随机域模型进行计算后的结果. (c) 图像正中给出了一个纹理的例子. (d) 图像的中心与图(c)相同, 四周镶嵌的是对(c)运用小波Markov随机域模型进行计算后的结果.

Markov随机域模型最早是由Cross和Jain [8] 提出的, 但是在D. Geman和S. Geman发表了对采样Gibbs分布的随机松弛算法的研究 [17] 后, 该模型才被计算和数学领域所关注. 为了简化叙述, 我们将讨论限制在一个随机向量 $X(n)$ 中, 其中 $n \in Z^d$ 在一个大小为 N 的网格 \mathcal{G} 中取值. 我们定义一个邻域系统 $\mathcal{N} = \{\mathcal{N}_n\}_{n \in \mathcal{G}}$ 使得 $n \notin \mathcal{N}_n$ 但 $m \in \mathcal{N}_n$ 成立的充要条件为: $n \in \mathcal{N}_m$. 对任意的 $\mathcal{G}_0 \subset \mathcal{G}$, 用 $X(\mathcal{G}_0)$ 表示 X 在 \mathcal{G}_0 上的值. 我们称 $p(X)$ 是关于 \mathcal{N} 的Markov随机域分布, 如果满足下式

$$p(X(n) | X(\mathcal{G} - \{n\})) = p(X(n) | X(\mathcal{N}_n)).$$

\mathcal{G} 的一个子集 C 被称为是集团 (clique), 如果 C 中的每一对元素都互为邻域. 用 \mathcal{C} 表示所有集团组成的集合. 如果 X 在有限个符号中取值, 那么根据Hammersley-Clifford定理可以证明 $p(X)$ 是一个Markov随机域当且仅当它可以被表示成关于 \mathcal{N} 的Gibbs分布

$$p(X) = \frac{1}{Z} \exp \left[- \sum_{c \in \mathcal{C}} \phi_c(X) \right],$$

其中 Z 是一个归一化常数, ϕ_{sc} 是只依赖于集团 C 中 X 的取值的势函数. Markov随机域理论在图像纹理的描述和图像恢复中有非常有趣的应用, 但由于对于合并相隔很远的象素有一定的困难, 因而其效果受到一定的影响. 有一些方法已经被用来解决这个问题, 包括重正化技术 [18] 在内.

Mumford和Zhu [35] 从另一个不同的观点, 他们不是从 X 的采样点, 而是从 X 的一个稀疏表示出发建立了Markov随机域模型. 用 $\mathcal{G} = \{g_\gamma\}_{\gamma \in \Gamma}$ 表示一本向量的“库”, 它可以是一组正交基, 也可以是更加冗余的基. 令 $X_\gamma = \langle X, g_\gamma \rangle$. 邻域系统 $\mathcal{N} = \{\mathcal{N}_\gamma\}_{\gamma \in \Gamma}$ 定义在 Γ 上. 例如, 如果 $\mathcal{G} = \{\psi_{j,n}^k\}_{k,j,n}$ 是一个二维小波基, 那么指标 $\gamma = (k, j, n)$ 说明了它的方向 k 、分辨率 2^j 和位置 $2^{-j}n$. 假定邻域 $\mathcal{N}_{(k,j,n)}$ 中还包含了满足 $|j-j'| \leq 1$, 位置 $2^{-j'}n'$ 与 $2^{-j}n$ 非常接近的小波 $\psi_{j',n'}^{k'}$. 多分辨率的特征允许我们可以建立包含有近程和远程相关性的Markov随机域模型.

为了从观察信号 $\{X_p^{obs}\}_{0 \leq p < P}$ 建立Markov随机域 X 的模型, 我们计算 M 个集团 $\{C_m\}_{0 \leq m < M}$ 上势函数 ϕ_{C_m} 的平均量值

$$\mu_{C_m}^{obs} = \frac{1}{P} \sum_{p=1}^P \phi_{C_m}(X_p^{obs}).$$

如果X是平稳的, 那么只要通过位置平移就可得到所有 ∂_{C_m} 空域均值. 这些经验平均值就是我们在模型X中对 $E\{\partial_{C_m}(X)\}$ 的估计值. 在很多情况下, 集团中只有两个元素 $C = \{\gamma, \gamma'\}$. 对应的协方差量值为 $\partial_C(X) = X_\gamma X_{\gamma'}$, 但不同的势函数如p阶矩(见下式)具有不同的用途.

$$\text{对 } p > 0, \quad \phi_C = |X_\gamma|^p |X_{\gamma'}|^p. \quad (14)$$

运用最大熵原理可以选择达到最大熵值的p(X)

$$p(X) = \operatorname{argmax} \{- \int p(X) \log p(X) dX\},$$

对应的约束条件是

$$\text{对 } 1 \leq m \leq M, \quad E\{\phi_{C_m}(X)\} = \int \phi_{C_m}(X) p(X) dX = \mu_{C_m}^{obs}. \quad (15)$$

为了使熵值达到极大, 对观察值 $\mu_{C_m}^{obs}$, 最终的p(X)应该是已给定先验知识的“最均匀”的分布. 因此它不再包含更多有用的“信息”. 可以通过拉格朗日乘数法求得解

$$P(X, \Lambda) = \frac{1}{Z(\Lambda)} \exp\left(- \sum_{m=1}^M \lambda_m \phi_{C_m}(X)\right). \quad (16)$$

如果势函数满足一个线性无关条件的话, 则其参数向量 $\Lambda = \{\lambda_m\}_{1 \leq m \leq M}$ 可被约束条件(15)唯一决定.

如果 $\phi_{C_m}(X)$ 表示协方差量值, 那么(16)式是Guass过程的概率分布, 并且如果 ϕ 是正交基, 那么 Λ 可以通过对带状协方差矩阵求逆进行计算. 求极大熵问题实际上是一个凸问题 [17], 但对于一般的势函数 ∂_{C_m} 不能用解析的方法求解向量 Λ . 可以通过数值计算方法对 $E_{p(X, \Lambda)}\{\partial_{C_m}(X)\}$ 进行递归来求解 Λ , 并按照梯度下降至满足条件(15)的情况下刷新 Λ . 需要指出, 估计 $E_{p(X, \Lambda)}\{\partial_{C_m}(X)\}$ 的方法是Gibbs采样器或者Monte Carlo马尔可夫链方法, 这在计算上是耗费巨大的.

Mumford和Zhu [35], 以及Simoncelli和Portilla [32], 用上面的Markov随机域理论构建了一个只用单幅图像纹理的模型. Simoncelli和Portilla [32] 的Markov模型是在小波基下计算的, 伴有对协方差量值和一阶矩的约束条件. 纹理图(7)a是在平稳性假设下用以计算模型中的参数 Λ 的单幅纹理图像. 图7(b)是用Markov模型经小波计算得到的一个结果, 它和原来的纹理非常接近, 主观视觉在小于0.1秒的时间内不能辨别出两者的差别. 图7(c)是“文字”信息作为纹理的类似的一个例子, 它同样可以在Markov模型下用小波方法进行计算. 图7(d)是通过把Markov模型的一个计算结果加到边缘地带后纹理图像, 而不是对中央区域的信息进行描述得到的.

Markov随机域理论为在合适的表达式上进行交互式稀疏处理提供了一个一般性框架来构造随机过程. 模型的有效性依赖于对表达式和势函数 ∂_C 的选择. 怎样去优化以上两个因素、以及分析函数空间上Markov随机域的性质依然是一个尚未解决的开放问题.

作者单位: 陈 刚(浙江大学高等数学研究所.)
张 亘(复旦大学中法应用数学中心.)
池哲儒(香港理工大学电子与资讯工程学系.)

注: *这是经Stéphane Mallat同意的其在1998年国际数学家大会上的专题报告的中译稿.

参考文献

- 1 L. Alvarez, F. Guichard, P. L. Lions and J. M. Morel, Axioms and fundamental equations of image processing. *Archieve for Rational Mechanic*, 123:199~257, 1993.
- 2 A. Benassi, S. Jaffard and D. Roux, Elliptic Gaussian random processes. *Revista Mathematical beroamericana*, 13:19~90, 1997.
- 3 S. Chen and D. Donoho, Atomic decomposition by basis pursuit. In *SPIE International Conference on Wavelets*, San Diego, July 1995.
- 4 A. Cohen, R. DeVore, P. Pertrushev, and H. Xu, Non-linear approximation and the space $BV(\mathbb{R}^2)$. *American J. of Math.*, to appear, 1998.
- 5 R. R. Coifman and Y. Meyer, Remarques sur l'analyse de Fourier a fenetre. *C. R. Acad. Sci.*, 259~261, 1991.
- 6 R. R. Coifman and M. V. Wickerhauser, Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, 38(2):713~718, 1992.
- 7 A. Croisier, D. Esteban, and C. Galand, Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *Int. Conf. on Inform. Sciences and Systems*, 443~446, Patras, Greece, August 1976.
- 8 G. Cross, and A. Jain, Markov random field texture models. *IEEE Trans. On PAMI*, 5:25~39, 1983.
- 9 I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.
- 10 I. Daubechies, *Wavelets and other phase space localization methods*. Proc. of ICM 94, Birkhauser Verlag, Switzerland, 1995.
- 11 J. G. Daugmann, Two-dimensional spectral analysis of cortical receptive field profile. *Vision Research*, 20:847~856, 1980.
- 12 R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decompositions. *Amer. J. Math.*, 114:737~785. 1992.
- 13 R. DeVore, *Nonlinear Approximation*. Acta Numerica, 51~150, Cambridge University Press, 1998.
- 14 D. Donoho and I. Johnstone, Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425~455, 1994.
- 15 D. Donoho, Abstract statistical estimation and modern harmonic analysis. Proc. of ICM 94, Birkhauser Verlag, Switzerland, 1995.
- 16 N. Dyn and S. Rippa, Data-dependent triangulations for scattered data interpolation and the finite element approximation. *Applied Num. Math.*, 12:89~105, 1993.
- 17 S. Geman, and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on PAMI*, 6:721~741, 1984.
- 18 B. Gidas, A renormalization group approach to image processing. *IEEE Trans. on PAMI*, 11:164~180, 1989.
- 19 S. Mallat, Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315:69~87, 1989.
- 20 S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397~3415, 1993.
- 21 S. Mallat, Z. Zhang, and G. Papanicolaou, Adaptive covariance estimation of locally stationary processes. *Annals of Stat.*, 26(1):1~47, 1997.
- 22 S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, Boston, MA, 1998.
- 23 H. S. Malvar, *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA, 1992.
- 24 D. Marr, *Vision*. W. H. Freeman and Co., San Fransisco, 1982.
- 25 Y. Meyer, *Wavelets and Operators*. Advanced Mathematics. Cambridge University Press, 1992.
- 26 D. Mumford, Mathematical Theories of Shape: Do they model perception? *SPIE: Geometric Methods in Computer Vision*, 1570:2~10, 1991.
- 27 J. F. Muzy, E. Bacry and A. Arneodo, The multifractal formalism revisited with wavelets. *Int. J. of Bifurcation and Chaos*, 4:245, 1994.
- 28 S. Osher, and L. Rudin, Feature-oriented image enhancement using shock filter. *SIAM J. on Numerical Analysis*, 27:919~940, 1990.
- 29 M. Priestley, Evolutionary spectra and non-stationary processes. *J. Roy. Stat. Soc., Ser. B*, 27:204~237, 1965.
- 30 J. M. Shapiro, Embedded image coding using zero-trees of wavelet coefficients. *IEEE Trans. Signal Proc.*, 41(12):3445~3462, 1993.

31 G. Shapiro, and A. Tannenbaum, On invariant curve evolution and image analysis. J. of Functional Analysis, 119(1):79~120, 1993.

32 E. Simoncelli and J. Portilla, Texture characterization via second-order statistics of wavelet coefficient amplitudes. Proc. of 5th IEEE Int. Conf. On Image Proc., Chicago, October 1998.

33 M. Vetterli and J. Kovacevic, Wavelets and Subband Coding. Prentice-Hall, Englewood Cliffs, NJ, 1995.

34 J. W. Woods and S. D. O'Neil, Sub-band coding of images. IEEE Trans. Acoust., Speech, and Signal Proc., 34(5):1278~1288, 1986.

35 S. C. Zhu, and D. Mumford, Prior learning and Gibbs reaction-diffusion. IEEE Trans. on PAMI, 19:1236~1250, 1997.

译文1999年2月25日收到.