

傅里叶变换：MP3、JPEG和Siri背后的数学

算法爱好者 2016-05-19

(点击上方公众号，可快速关注)

英文：Aatish Bhatia

译者：伯乐在线 - programmer_lin

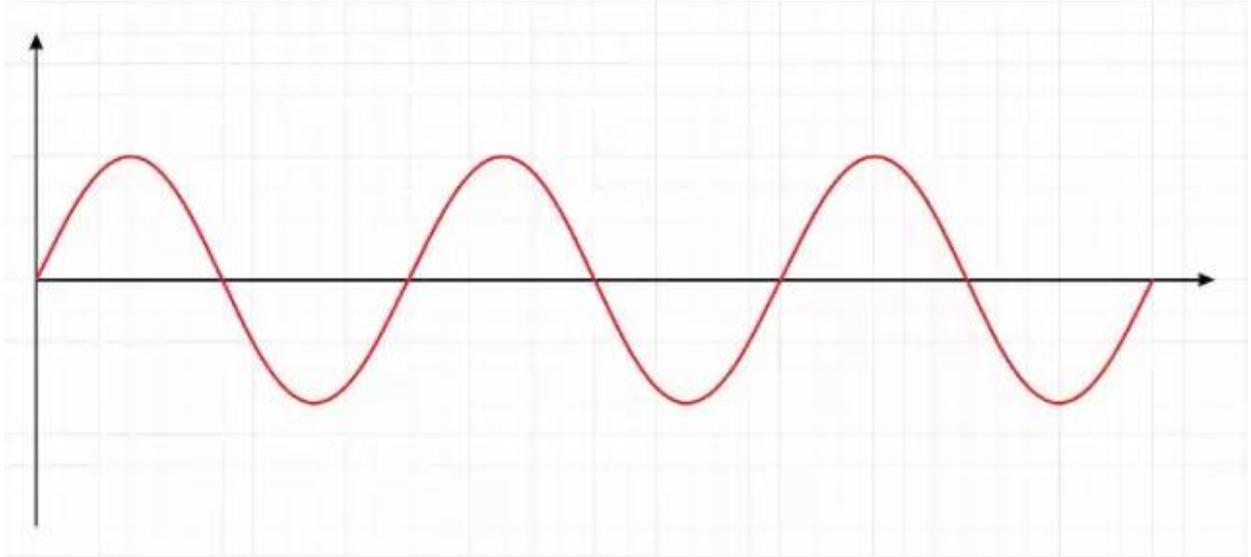
链接：<http://blog.jobbole.com/51301/>

九年前，当我还坐在学校的物理数学课的课堂里时，我的老师为我们讲授了一种新方法，给我留下了深刻印象。

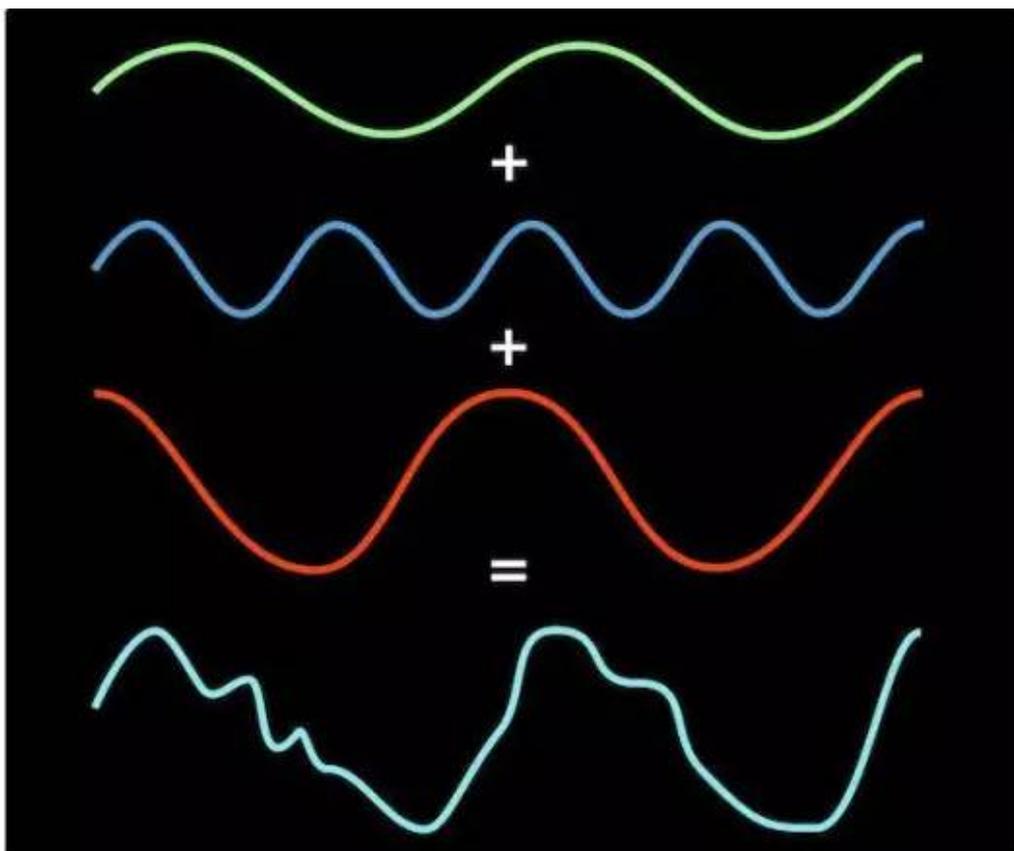
我认为，毫不夸张地说，这是对数学理论发现最广泛的应用。应用的领域包括：量子物理、射电天文学、MP3和JPEG压缩、X-射线晶体学、语音识别、PET或MRI扫描。这种数学方法叫做傅里叶变换，这种方法因18世纪的法国物理学家、数学家约瑟夫·傅立叶（Joseph Fourier）而得名。这种方法甚至被詹姆斯·沃森和弗朗西斯·克里克用来解码由罗莎琳德·富兰克林通过X射线得到的DNA双螺旋结构。（克里克是傅里叶变换的专家，他写过一篇名为《傅里叶变换在观鸟者中的应用》的趣文，来向名为沃森的观鸟爱好者解释这一数学概念。）

无论你在听MP3格式的歌曲，还是在网页上浏览图片，或者向SIRI提问，甚至打开收音机时，你都可能在日常生活中应用了演化的傅里叶变换。（顺便说一下，傅里叶并不是一个敷衍取巧的人。在他研究的理论物理和数学之外，他还是第一个发现温室效应的人<PDF>）。

那么，什么是傅里叶理论呢，为什么他的这个理论如此有用？想想你在钢琴键盘上敲响一个音符。当你按下琴键的时候，钢琴中有一个小锤来来回回地敲击一根琴弦（对于音准do大约是440次每秒）。随着琴弦振动，它周围的空气分子也来回震动，从而创造了一波震动的空气分子，我们称之为声。如果你能看空气中进行的这种有规律的舞蹈，你会发现一系列平稳，起伏的，无休止的重复。这就是所谓的正弦波曲线，或正弦波。（特别说明：在钢琴的例子中，肯能会产生不止一条正弦波实际演奏中，钢琴音色的丰富性正是来源于在主要正弦波之外的那些轻柔的泛音。钢琴的音符可以大致模仿一条正弦波，但是对于单一的正弦波声音来说，音叉发出的声音是一个更加贴切的例子。）



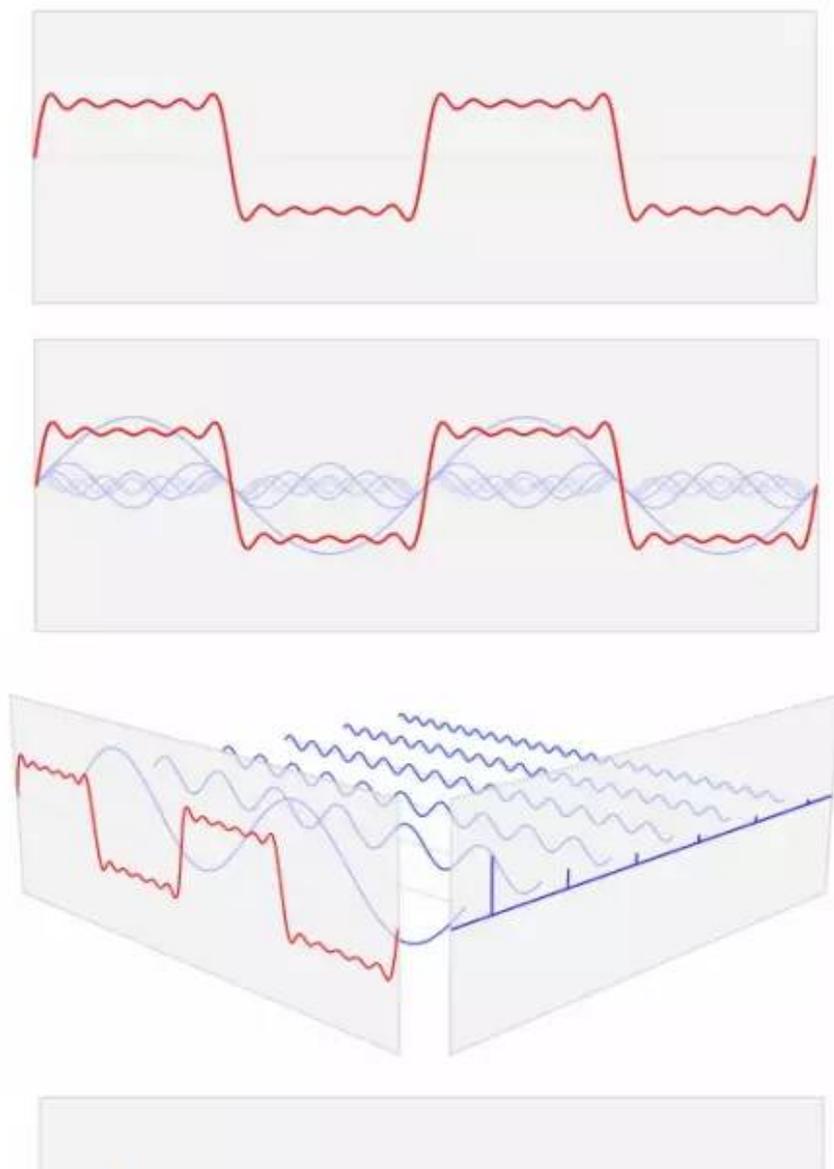
现在，让我们暂时放下单独一个音符，而考虑由三个按键同时发出的和弦声。和弦结果的声波并不漂亮——它看起来杂乱无章。但是，在这混乱的背后有一个简单的模型。毕竟，和弦只是三个音符的相互融合与碰撞，因此这样混乱的声波，实际上只是三种音符（正弦曲线）的和而已。

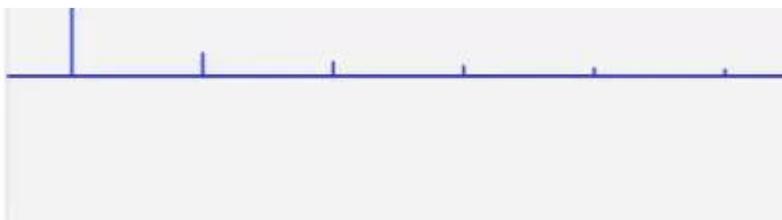


傅里叶认为这不仅仅音乐和弦的特殊属性，而可以推广应用到一切重复的波形中，无论这个波形是方形，圆形，波浪，三角形伙食其他。傅里叶变换像是一种数学棱镜——你输入一个波形并且将这种波形分解为不同成分——这些音符（正弦曲线）会相互叠加而形成新的重建波形。

如果这听起来有一些抽象的话，有一些可视化的方式来使得傅里叶的方法更加直观。第一种方法是有卢卡斯（Lucas V. Barbosa）提出的。他是一名来自巴西的物理学学生，他将业余时间无偿用于为维基百科制作关于数学和科学的动画，在维基百科上他被称为“LucasVB”。

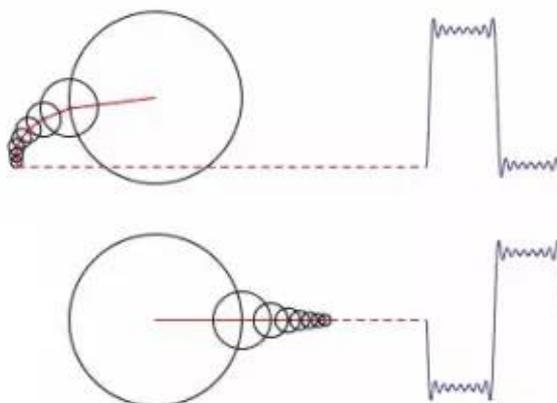
那么，现在让我们来看看输入一个方形波，经过傅里叶变换后，会输出怎样的波形。





在这些图形中（[点击这里可查看 GIF 动态图](#)），红色的方形波被分离为单纯波形的集合（蓝色的正弦曲线）。将这些蓝色波形认为是红色波形的数学成分列表。在这个比喻中，傅里叶变换就像是一种药方——他准确地告诉你要重建原始波形，每一种简单波形你要使用多少。动画中的垂直蓝色线，是每种波形数量的直观表示。

思考这个问题，还有一种由马修·亨德森（Matthew Henderson or Matthen）提供的方法。他是剑桥大学的博士生，并且对于创建数学动画模型非常有兴趣。他用圆形而不是正弦曲线来解释傅里叶变换。这种方法包含了一组不同大小的源，每个圆的圆心都在一个更大的圆的边缘上。然后，这些圆开始转动，大圆在小圆的周围摆动，小圆的运动速度大于大圆。如果你追踪最小圆上一个点的运动轨迹，如下面的动画和截图所示，你可以重建任意形状的波形。傅里叶变换再一次告诉你波形是怎样产生的：以怎样的速度去移动哪些圆。



如果你年纪够大以至于你用过呼吸量描记仪，那么通过层叠的齿轮来描述复杂模型的想法你可能很熟悉。LucasVB在同一个动画上制作了互动的版本，使得你可以随意改变圆圈的大小。

总的来说，傅里叶变换告诉你，在一个整体的波形中，每一个单独的“音符”（正

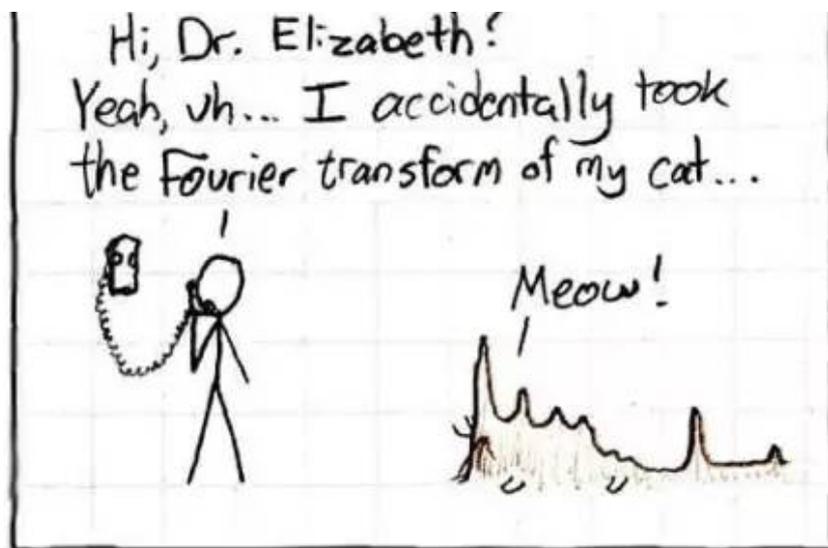
弦曲线或是圆圈)的比例。这就是傅里叶变换如此有用的原因。想象一下,你正在和你的朋友通电话,同时你想让他们能够画出近似方形的波形。复杂的方式是读出一长串的数字,每个数字表示了相应时间点上波形的高度。有了这些数字,你的朋友可以耐心地绘制出原始波形。这就是原始的音频格式比如WAV的基本原理。但是,如果你的朋友知道傅里叶变换,那么你可以更加聪明地完成这个工作:你只要告诉他们少量数字——上图中提到的不同的圆的尺寸。他们可以用这些圆来重建原始波形。

这不仅仅是数学花招。傅里叶变换出现在几乎所有存在波形的地方。无处不在的MP3格式使用一种变形的傅里叶变换来达到相比之前的WAV(读作“wave”)更大的压缩率。对于每个音频片段,傅里叶变换将音频波形分解为它的成分音符并且保存下来,从而代替存储原始波形。傅里叶变换还可以告诉你在一首歌中每个音符所占的比例,你可以知道哪些音符是这首歌的基本元素。音调很高的音符并不重要(我们的耳朵几乎不能听见),因此,MP3格式放弃保存这些音符,从而取得了更高的数据压缩率。这正是高保真音响爱好者不喜欢MP3格式的原因——它不是一种无损的音频格式,高保真爱好者表示他们可以听出其中的差别。

这也是智能手机的应用程序Shazam怎样识别一首歌的原理。它将音乐分割成块,利用傅里叶变换算出每一块中的音符成分。然后它搜索数据库,来寻找这样的“音符指纹”与他们已有文件中的一首歌相匹配。语音识别同样使用“傅里叶——指纹”的思想,将你的声音与已知单词列表进行比较。

你也可以在图像上利用傅里叶变换。有一个极好的视频来说明你怎么利用圆圈来绘制辛普森的脸。在线百科全书Wolfram Alpha采用了相似的理念来绘制名人头像。听起来,这似乎可以存下来用于一个恶搞的鸡尾酒会,但是,这种方法也用于将图像压缩为JPEG文件。在以前的微软绘图中,图像是用位图(BMP)存放的,这种文件包含了一长串的数字,代表对每个像素点的颜色编码。JPEG格式就相当于图像格式中的MP3格式。建立一个JPEG文件,你首先将图片分割为很小的块,每块都是8像素*8像素。对于每个像素块,你可以用与重建辛普森的脸相同的画圆的办法来重建局部图像。正如MP3放弃保存高音一样,JPEG不保存极小的圆。这样做的结果是:牺牲了小部分的画面质量,来取得文件大小的巨大压缩。这样的理念,使得我们都喜欢的可视化网络世界成为可能(同时最终让我们得到了GIF格式)。





在科学研究中，傅里叶变换又有怎样的应用呢？我在推特上邀请科学家们来描述他们在工作中是怎样应用傅里叶的思想的。他们的回复使我惊讶。做出回复的科学家表示，他们正在利用傅里叶变换：研究不同的潜水器结构与水流的相互作用，试图预测即将到来的地震，识别距离遥远的星系的组成部分，寻找热量大爆炸残余物中的新物理成分，从x射线衍射模式揭示蛋白质的结构，为NASA分析数字信号，研究乐器的声学原理，改进水循环的模型，寻找脉冲星（自转的中子星），用核磁共振研究分子结构。傅里叶变换已经被用于通过破译油画中的化学物质，来识别假冒的杰克逊·波洛克绘画。

哇！这仅仅是一个相当传统的小数学技巧！

算法爱好者

专注算法相关内容



微信号: AlgorithmFans



长按识别二维码关注

伯乐在线 旗下微信公众号

商务合作QQ: 2302462408