arXiv:2302.08721v1 [physics.data-an] 17 Feb 2023

加性与非加性光谱背景校正中的插值逆离散小波变换

TEEMU HÄRKÖNEN[1,*] 和 ERIK VARTIAINEN[1]

[1] 工程科学学院，拉彭兰塔-拉赫蒂理工大学，*Yliopistonkatu 34*，拉彭兰塔，*FI-53850*，芬兰[*]
teemu.harkonen@lut.fi

摘要： 我们证明了使用插值逆离散小波变换作为一种通用工具，对光谱中的加性或乘性背景或误差信号进行建模的适用性。此外，我们提出了一种无监督方法，用于估计最优小波基及模型参数。我们将该方法应用于酞菁蓝、苯胺黑、萘酚红、颜料黄150和颜料红264颜料的实验拉曼光谱，以去除其加性背景；并应用于腺苷磷酸、果糖、葡萄糖和蔗糖的相干反斯托克斯拉曼散射光谱，以去除其乘性背景信号。

1. 引言

拉曼光谱学是进行分子鉴定与分析的一项强大技术。由于拉曼光谱学及其变体，如相干反斯托克斯拉曼散射（CARS）、表面增强拉曼光谱（SERS）或表面增强共振拉曼光谱（SERRS），均具有无创且无损的特点，因此除其他众多应用外，它们尤其适用于生物医学领域 [1,2]。

然而，光谱通常伴有某种形式的背景信号，从而掩盖了由分子振动产生的拉曼信号。背景校正方法包括非对称最小二乘法 [3,4], 多项式拟合 [5], 和基于小波的方法 [6,7]。更全面的综述请参见 [8]。

本研究重点探讨此前在 [9]中提出的插值离散小波变换，它是小波棱镜法 [10] 的扩展，用于校正相干反斯托克斯拉曼散射的乘性误差。我们证明，对小波重构层级进行插值，是一种可行的建模方法，可用于描述一般光谱的背景过程；此外，我们还提出了一项准则，用于自动选择小波基及相应的模型参数。该模型仅由一个或两个参数进行参数化，因此可应用于高维参数可能难以处理的情形，例如贝叶斯推断，参见例如 [9,11]

本文其余部分的结构如下。第 2 节构建用于加性和乘性背景的插值小波变换模型，并介绍优化方案。第 3 节记录实验细节，第 4 节展示实验拉曼光谱和相干反斯托克斯拉曼散射光谱的结果。最后，第 5 节总结本研究。

2. 插值小波背景模型

我们将测得的光谱建模为含加性背景 $B(\nu_k,\ c,\ p,\ \mathrm{M})$

$$y(\nu_k) = f(\nu_k) + B(\nu_k, c, p, \mathcal{M}), \tag{1}$$

其中，ν_k 表示测量位置，c 表示背景的恒定偏移，p 表示插值小波细节层级，M 表示所选的小波基，可将其视为一个模型选择问题。我们按照 [9], 中的方式对背景建模，并在以下两者之间进行插值

离散小波重构层级 $D_j(\nu)$ 定义为

$$B(\nu_k, c, p, \mathcal{M}) = c + \sum_{j=\lceil p+1 \rceil}^{J} D_j(\nu, \mathcal{M}) + (1-\beta) D_{\lceil p \rceil}(\nu, \mathcal{M}), \tag{2}$$

其中， J 表示所选小波分解层级的最大值， $p \in [1, J]$、 $\beta = p - \lfloor p \rfloor$，离散小波变换（DWT）根据数据 $y(\boldsymbol{\nu}) = (y(\nu_1)$、.....、 $y(\nu_K))^T$ 使用基 M_i 计算。关于非插值方法，参见 [10]。

对于乘性光谱，例如相干反斯托克斯拉曼散射的情况，测量光谱可以建模为

$$z(\nu_k) = S(\nu_k)\varepsilon_{\mathrm{m}}(\nu_k; c, p, \mathcal{M}), \tag{3}$$

其中， $z(\nu_k)$ 是在 ν_k 处的测量值， $S(\nu_k)$ 表示无误差光谱，而 $\varepsilon_{\mathrm{m}}(\nu_k;\ c, p)$ 是调制背景或误差函数。通过取对数，可以将式 (3) 中的表达式转换为加性形式：

$$\log z(\nu_k) = \log S(\nu_k) + \log \varepsilon_{\mathrm{m}}(\nu_k; c, p, \mathcal{M}), \tag{4}$$

这样便可以如上所述对调制误差函数进行建模：

$$\log \varepsilon_{\mathrm{m}}(\nu; c, p, \mathcal{M}) = c + \sum_{j=\lceil p+1 \rceil}^{J} D_j(\nu, \mathcal{M}) + (1-\beta) D_{\lceil p \rceil}(\nu, \mathcal{M}), \tag{5}$$

其中，离散小波变换根据对数数据计算，即 $\log\ z(\boldsymbol{\nu}) = (\log\ z(\nu_1), \ldots, \log\ z(\nu_K))^T$。我们优化参数 p，并按如下方式建立模型 M。对于加性拉曼光谱，使目标函数最小的参数 $\hat{}p$ 和模型 $\hat{}\mathrm{M}$ 为

$$(\hat{p}, \hat{\mathcal{M}}) = \underset{(p,\mathcal{M})}{\arg\min} \left\| f(\boldsymbol{\nu}) - \min f(\boldsymbol{\nu}) \right\|_1 = \underset{(p,\mathcal{M})}{\arg\min} \sum_{k=1}^{K} \left| f(\nu_k) - \min f(\boldsymbol{\nu}) \right|, \quad p < p_{\max} \tag{6}$$

其中，我们对背景的恒定偏移 c 进行建模，使光谱的最小值为零，即 $f(\boldsymbol{\nu})$ –min $f(\boldsymbol{\nu})$； $p_{\max}$ 是为 p 选定的最大界限，而 $\|\cdot\|_1$ 表示 L^1 范数。对于相干反斯托克斯拉曼散射光谱，我们采用类似的方案：

$$(\hat{p}, \hat{\mathcal{M}}) = \underset{(p,\mathcal{M})}{\arg\min} \left\| \mathrm{Im}\left\{\chi_3(\boldsymbol{\nu})\right\} - \min \mathrm{Im}\left\{\chi_3(\boldsymbol{\nu})\right\} \right\|_1, \quad p < p_{\max} \tag{7}$$

在这里我们可以设置 $c = 0$，例如参见 [10],，而 Im $\{\chi_3(\boldsymbol{\nu})\}$ 是从校正光谱 $S(\boldsymbol{\nu})$ 中提取的拉曼光谱。有关 Im $\{\chi_3(\boldsymbol{\nu})\}$ 的提取详情，请参见 [12–15]。

3. 实验细节

在接下来的章节中，我们将介绍数值示例所用的实验样品。

拉曼样品

这些拉曼光谱取自一个免费的在线数据库，该数据库收录了现代及当代艺术中所用颜料的拉曼光谱（The standard Pigments Checker v.5） [16]。此处使用的所有光谱都受到荧光干扰，导致拉曼线形中出现较大的背景成分。

相干反斯托克斯拉曼散射样品

多路复用相干反斯托克斯拉曼散射光谱学中使用的糖样品，是 D-果糖、D-葡萄糖及其二糖组合蔗糖的等摩尔水溶液（α-D-吡喃葡萄糖基-（{v2）-β-D-呋喃果糖苷）。制备样品时，将糖样品溶于缓冲溶液（50 mM HEPES，pH=7），使其摩尔浓度均为 500 mM [17]。腺苷磷酸样品是腺苷一磷酸、二磷酸和三磷酸（AMP、ADP、ATP）在水中的等摩尔混合物，总浓度为 500 mM [18]。对于 AMP、ADP 或 ATP，腺嘌呤环振动 [19] 均出现在相同频率附近，即约 1350 cm^{-1}。900 至 1100 cm^{-1} 之间的磷酸基振动可用于区分不同的核苷酸 [20]。ATP 的三磷酸基团在 1123 cm^{-1}处显示出强共振，而 AMP 的一磷酸基团共振出现在 979 cm^{-1} 处。ADP 则在两者之间的 1100 cm^{-1} 处出现展宽的共振。

多路复用相干反斯托克斯拉曼散射光谱学

所有用于验证我们方法学的相干反斯托克斯拉曼散射光谱均采用多路复用相干反斯托克斯拉曼散射光谱仪记录，关于该仪器的详细描述可见其他文献 [20,21]。简而言之，一台10 ps和一台80 fs的锁模钛蓝宝石激光器经电子同步后，分别在多路复用相干反斯托克斯拉曼散射过程中提供窄带泵浦/探测激光脉冲和宽带斯托克斯激光脉冲。泵浦/探测和斯托克斯脉冲的中心波长为710 nm。斯托克斯激光器可在750至950 nm之间调谐。糖类光谱的探测波数范围为700至1250 cm^{-1}，AMP/ADP/ATP光谱的探测范围为900至1700 cm^{-1}。线偏振且彼此平行的泵浦/探测光束和斯托克斯光束被调整为共线，并通过消色差透镜聚焦进入串联比色皿。后者可沿垂直于光轴的方向平移，以便在其两个腔室中的任一个内进行测量，从而在近乎相同的实验条件下获得样品和非共振参照物的多路复用相干反斯托克斯拉曼散射光谱。样品处使用的典型平均功率中，泵浦/探测激光器和斯托克斯激光器分别为95 mW（对于AMT/ADP/ATP，为75 mW）和25 mW（105 mW）。反斯托克斯信号在前向散射几何构型中由第二个消色差透镜收集并准直，经短通滤光片和陷波滤光片进行光谱滤波后，聚焦进入配备CCD相机的光谱仪。每条相干反斯托克斯拉曼散射光谱的采集时间为：糖类光谱200 ms，AMP/ADP/ATP光谱800 ms。

4. 结果与讨论

我们将该方法应用于酞菁蓝、苯胺黑和萘酚红的实验拉曼光谱，以去除加性背景；并应用于腺苷磷酸、果糖、葡萄糖和蔗糖的实验相干反斯托克斯拉曼散射光谱，以进行乘性背景校正。结果分别见图1–9，其中展示了实验测量结果、估计的加性或乘性背景以及校正光谱。

对于小波模型 M，使用了 Daubechies 5–35、小波 symlet 5–25 和 Coiflet 1–5。总计涉及 57 种不同的小波基。

酞菁蓝、苯胺黑和萘酚红的结果显示，拉曼光谱的背景较为平滑，但萘酚红背景存在一些边界效应。颜料红264的结果中也能观察到类似的边界效应，见图5。图4展示了颜料黄150拉曼光谱的背景去除结果。较小且实际上近乎恒定的背景被成功去除，未观察到过拟合。针对酞菁蓝、苯胺黑、萘酚红、颜料黄150和颜料红264拉曼光谱优化得到的小波基依次为 symlet 5、symlet、symlet 5、Daubechies 22 和 symlet 11。

对于相干反斯托克斯拉曼散射光谱，结果与 [9] 中所呈现的结果相似。图 6–8 中腺苷磷酸、果糖和葡萄糖的相干反斯托克斯拉曼散射光谱结果显示出不同的

缓慢变化的背景信号，其频率成分大致相似；这一点也可从得到的波小波插值参数（其数值大小相近）近似推断，分别为 $\hat{p} = 12$。652，13。402，13。616。图 9 中的蔗糖相干反斯托克斯拉曼散射结果呈现出更加线性的背景，这一点也体现在较低的小波插值参数值 $\hat{p} = 11$。702 上，表明与另外三种情况相比，其频率分量较低。腺苷磷酸、果糖、葡萄糖和蔗糖相干反斯托克斯拉曼散射光谱的优化小波基分别为 Daubechies 38、symlet 12、symlet 14 和 Daubechies 31。

小波背景模型表现出灵活性。对于图 1 中的酞菁蓝拉曼样品，估计背景实际上是恒定的。在图 5 和图 9 中，对于颜料红264和蔗糖，模型的表现类似于线性拟合。对于酞菁蓝、苯胺黑和萘酚红拉曼样品，以及腺苷磷酸、果糖和葡萄糖相干反斯托克斯拉曼散射样品，该模型能够捕捉不同程度的振荡特征。

拉曼

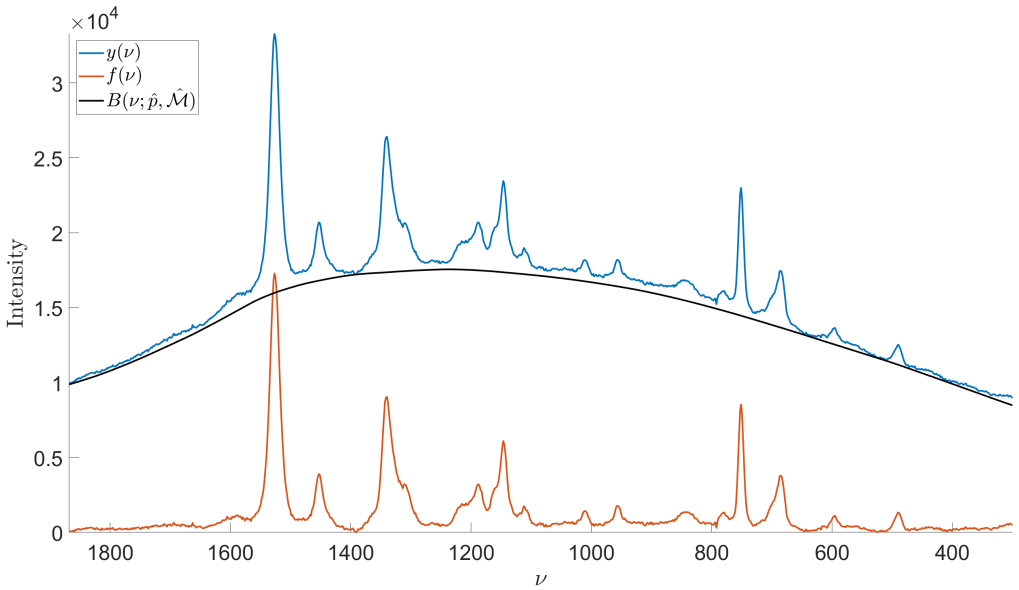


图 1. 蓝色为酞菁蓝的数据。红色为校正光谱 $f(\nu)$，黑色为估计的加性背景。所得背景采用 ˆM = Symlet 5 小波和 ˆ$p = 10$。663 获得。

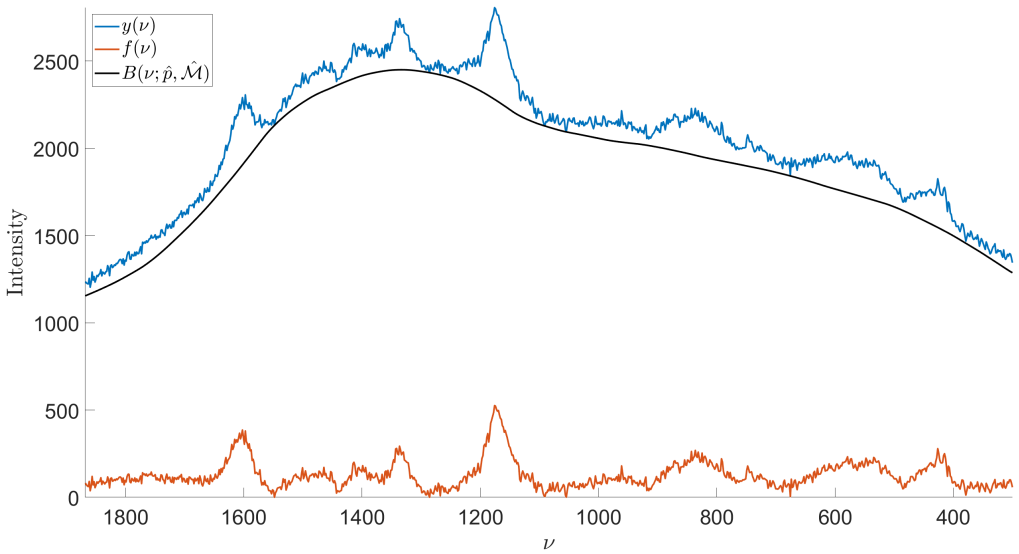


图 2. 蓝色为苯胺黑的数据。红色为校正光谱 $f(\nu)$，黑色为估计的加性背景。所得背景采用 ˆM = Symlet 5 小波和 ˆ$p = 11$。882 获得。

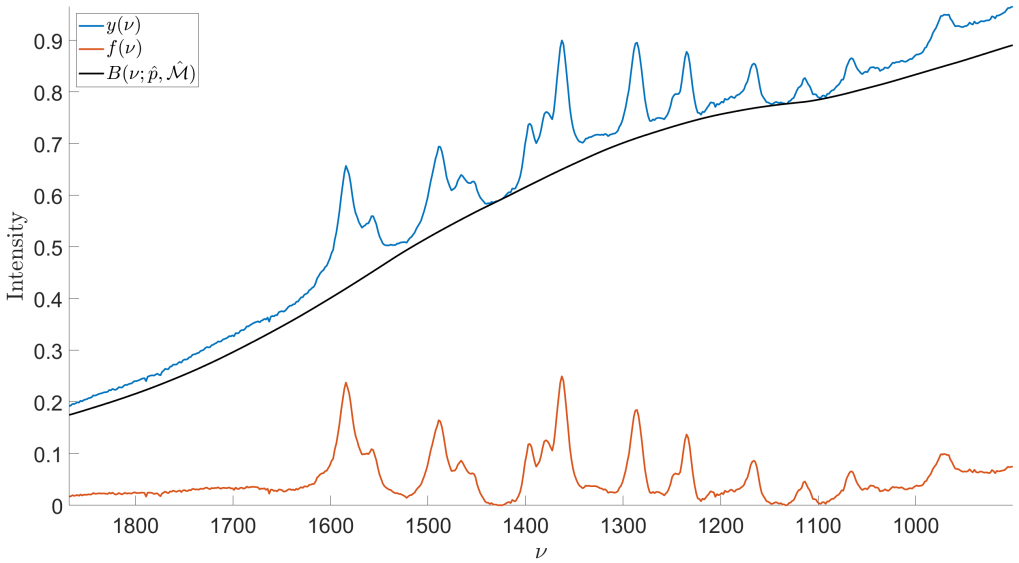


图3. 蓝色为萘酚红的数据。红色为校正光谱 $f(\nu)$，黑色为估计的加性背景。所得背景使用 $\hat{\mathcal{M}}$ = Symlet 5 小波和 $\hat{p} = 11$。591 得到。

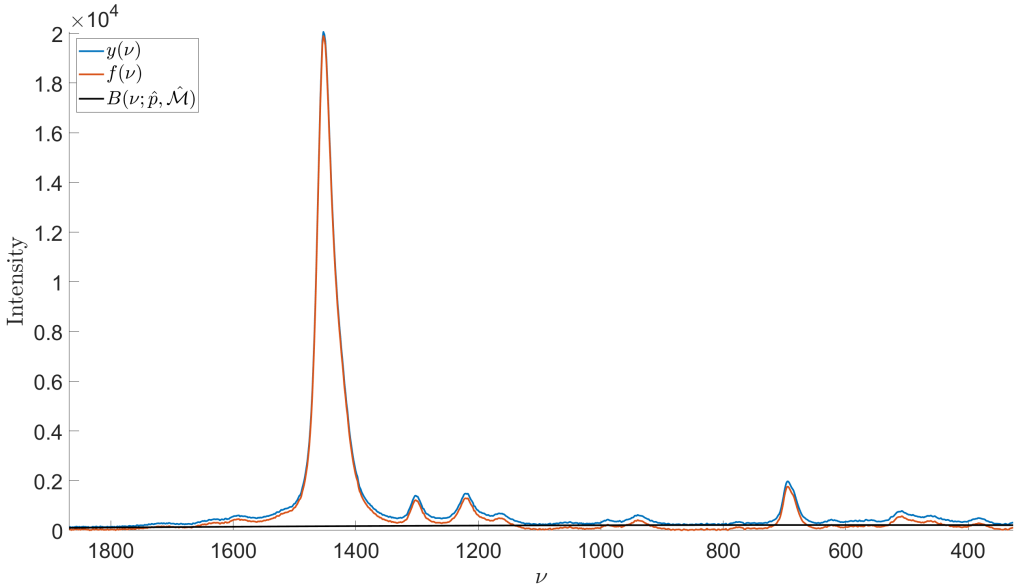


图4. 蓝色为颜料黄150的数据。红色为校正光谱 $f(\nu)$，黑色为估计的加性背景。所得背景使用 $\hat{\mathcal{M}}$ = Daubechies 22小波和 $\hat{p} = 7$。756 得到。

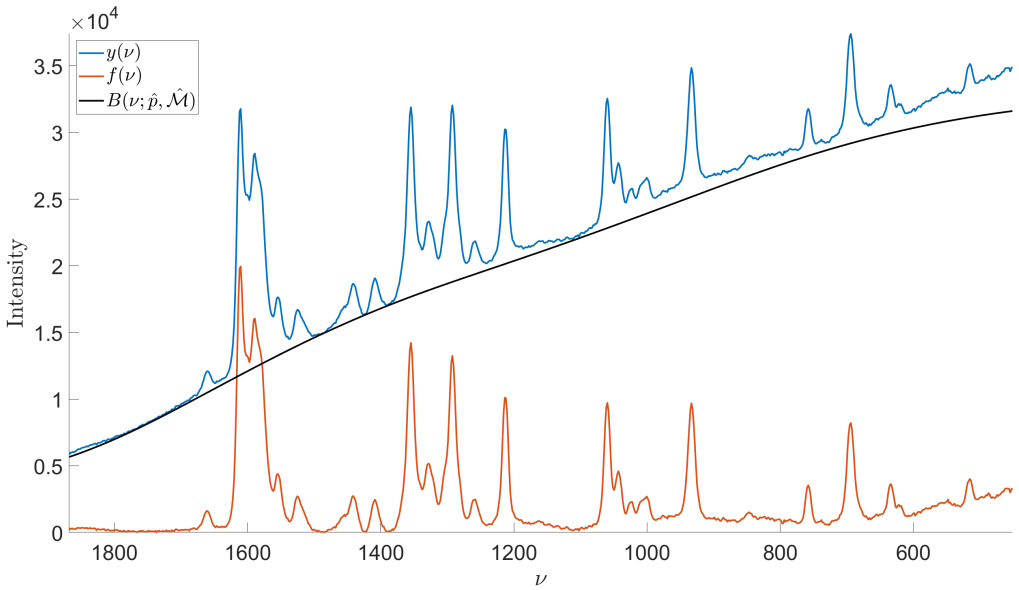


图5. 蓝色表示颜料红264的数据。红色表示校正光谱 $f(\nu)$，黑色表示估计的加性背景。所得背景是使用 ˆM = Symlet 11 小波并取 ˆp = 10。555 得到的。

CARS

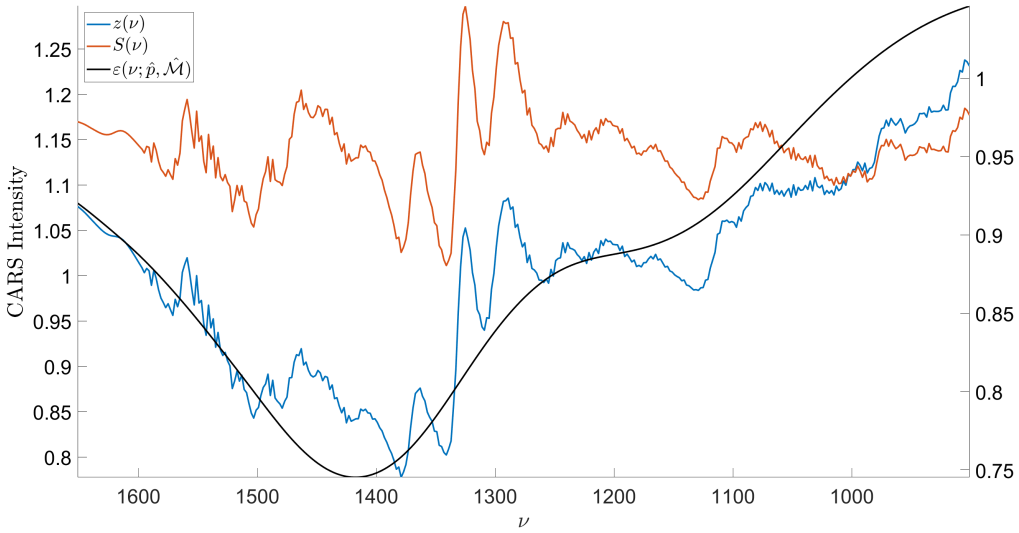


图 6. 蓝色为测得的腺苷磷酸相干反斯托克斯拉曼散射光谱。红色为校正光谱 $S(\nu)$，黑色为估计的乘性背景。所得背景使用 ˆM = Daubechies 10 小波和 ˆp = 12.652 得到。

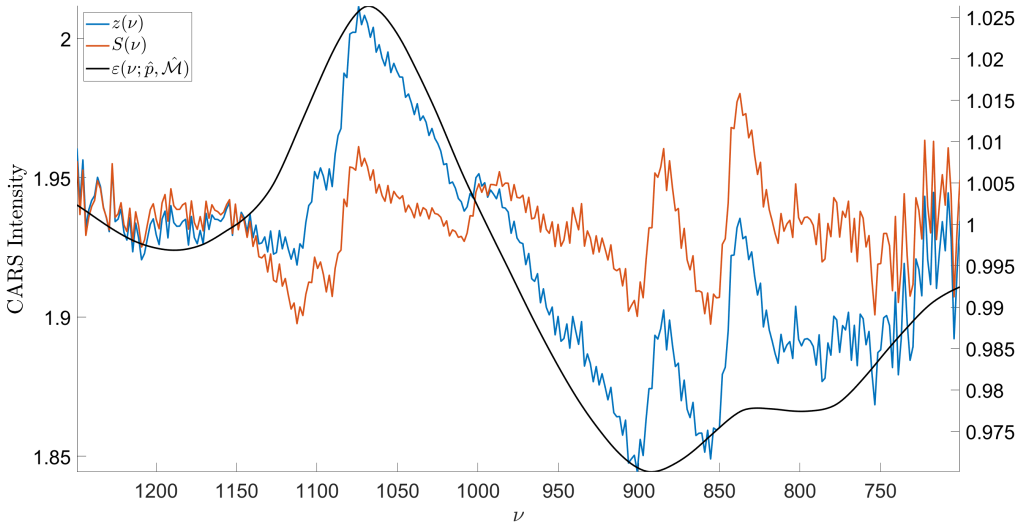


图 7. 蓝色为测得的果糖相干反斯托克斯拉曼散射光谱。红色为校正光谱 $S(\nu)$，黑色为估计的乘性背景。所得背景使用 ˆM = coiflet 3 小波和 ˆp = 13.402 得到。

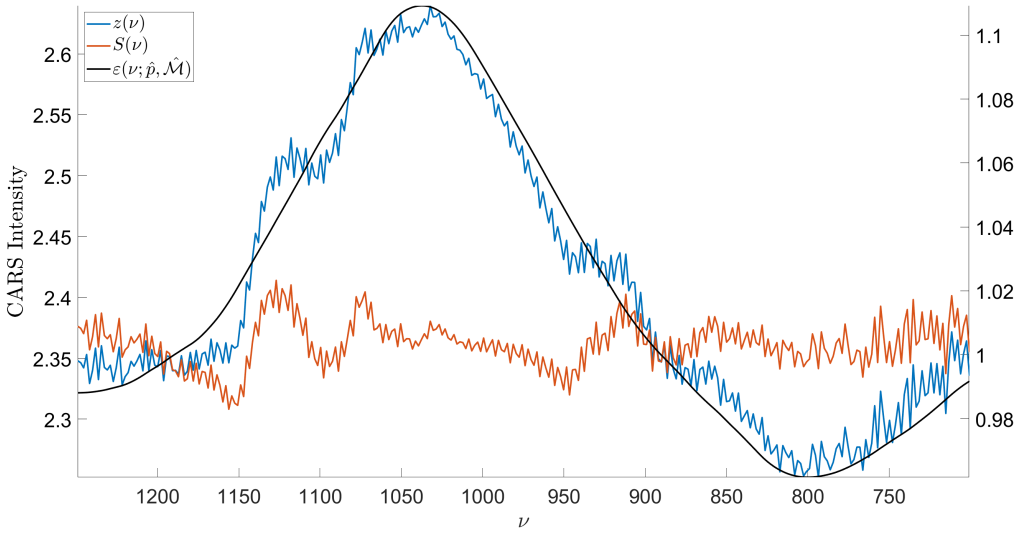


图 8. 蓝色为测得的葡萄糖相干反斯托克斯拉曼散射光谱。红色为校正光谱 $S(\nu)$，黑色为估计的乘性背景。所得背景采用 ˆM = Daubechies 5 小波和 ˆp = 13.616 得到。

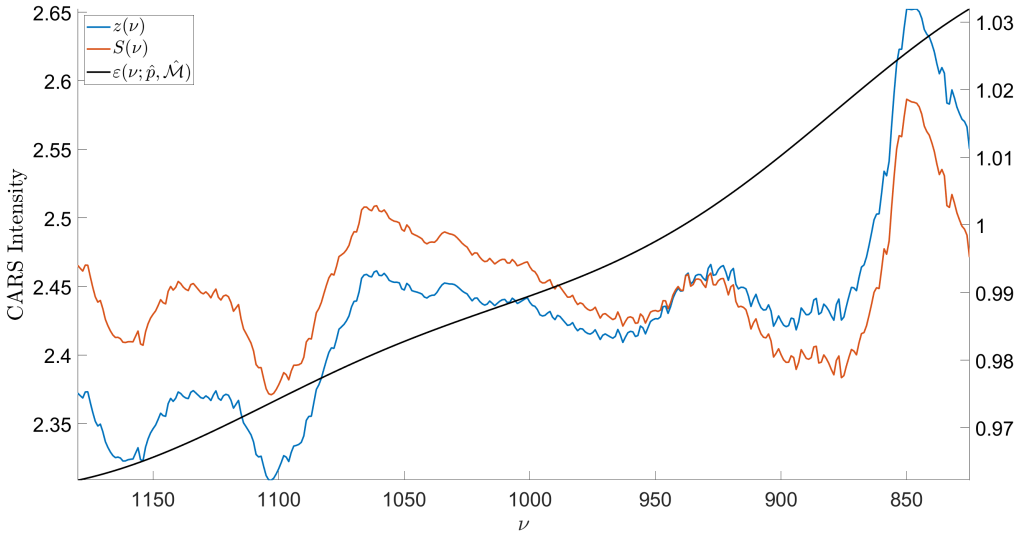


图 9. 蓝色为测得的蔗糖相干反斯托克斯拉曼散射光谱。红色为校正光谱 $S(\nu)$，黑色为估计的乘性背景。所得背景采用 ˆM = Daubechies 35 小波和 ˆp = 11.702 得到。

5. 结论

本研究提出了一种插值小波背景模型，用于对光谱的背景函数进行建模。我们针对拉曼测量和相干反斯托克斯拉曼散射测量，采用基于最小化校正后拉曼光谱面积的优化方案，成功优化了背景参数和小波基。该模型能够对一类灵活的背景函数进行建模，同时只需两个参数即可表征具有加性背景的拉曼光谱，只需一个参数即可表征具有乘性背景的相干反斯托克斯拉曼散射光谱。我们将所提出的方法应用于苯胺黑和酞菁蓝的实验拉曼光谱，以及腺苷磷酸、果糖、葡萄糖和蔗糖的实验相干反斯托克斯拉曼散射光谱。相干反斯托克斯拉曼散射光谱的结果与先前获得的结果一致。

经费支持。 作者获得了芬兰科学院的资助（项目编号：327734）。

利益披露。 作者声明不存在利益冲突。

数据可用性。 本文所述结果所依据的数据和软件可在通讯作者的 GitHub 个人代码库中公开获取，也可在提出合理请求后向作者索取。

参考文献。

1 CMcNay、D. Eustace、W. E. Smith、K. Faulds 和 D. Graham，“表面增强拉曼散射（SERS）和表面增强共振拉曼散射（SERRS）：应用综述”，Appl. Spectrosc. **65**，825–837（2011）。2. S. Li、Y. Li、R. Yi、L. Liu 和 J. Qu，“相干反斯托克斯拉曼散射显微术及其应用”，Front. Phys. **8**（2020）。3. H. F. M. Boelens、P. H. C. Eilers 和 T. Hankemeier，“符号约束有助于检测复杂光谱数据集之间的差异：以液相色谱-红外光谱为例”，Anal. Chem. **77**，7998–8007（2005）。4. S. He、W. Zhang、L. Liu、Y. Huang、J. He、W. Xie、P. Wu 和 C. Du，“采用改进的非对称最小二乘法对拉曼光谱进行基线校正”，Anal. Methods **6**，4402–4407（2014）。5. F. Gan、G. Ruan 和 J. Mo，“利用改进的迭代多项式拟合与自动阈值进行基线校正”，Chemom. Intell. Lab. Syst. **82**，59–65（2006）。6. C. M. Galloway、E. C. L. Ru 和 P. G. Etchegoin，“利用小波变换通过迭代算法去除光谱背景”，Appl. Spectrosc. **63**，1370–1376（2009）。7. M. Chi、X. Han、Y. Xu、Y. Wang、F. Shu、W. Zhou 和 Y. Wu，“一种基于小波变换的改进拉曼光谱背景校正算法”，Appl. Spectrosc. **73**，78–87（2019）。PMID：30251877。8. K. H. Liland、T. Almøy 和 B.-H. Mevik，“光谱多变量校准中基线校正方法的最佳选择”，Appl. Spectrosc. **64**，1007–1016（2010）。9. T. Härkönen、L. Roininen、M. T. Moores 和 E. M. Vartiainen，“相干反斯托克斯拉曼散射光谱学的贝叶斯定量”，The J. Phys. Chem. B **124**，7005–7012（2020）。10. Y. Kan、L. Lensu、G. Hehl、A. Volkmer 和 E. M. Vartiainen，“小波棱镜分解分析在相干反斯托克斯拉曼散射光谱学中的应用：准确、定量提取共振振动响应的工具”，Opt. Express **24**，11905–11916（2016）。11. M. Moores、K. Gracie、J. Carson、K. Faulds、D. Graham 和 M. Girolami，“拉曼光谱学的贝叶斯建模与定量”，arXiv（2016）。12. E. M. Vartiainen，“用于相干反斯托克斯拉曼散射光谱分析的相位恢复方法”，J. Opt. Soc. Am. B **9**，1209–1214（1992）。13. E. M. Vartiainen、H. A. Rinia、M. Müller 和 M. Bonn，“从拥挤的相干反斯托克斯拉曼散射光谱中直接提取拉曼线型”，Opt. Express **14**，3622–3630（2006）。14. Y. Liu、Y. J. Lee 和 M. T. Cicerone，“利用时域克拉默斯–克罗尼希变换进行宽带相干反斯托克斯拉曼散射光谱相位恢复”，Opt. Lett. **34**，1363–1365（2009）。15. M. T. Cicerone、K. A. Aamer、Y. J. Lee 和 E. Vartiainen，“最大熵与时域克拉默斯–克罗尼希相位恢复方法在功能上等效于相干反斯托克斯拉曼散射显微光谱学”，J. Raman Spectrosc. **43**，637 643 2012–（）。16.“The standard Pigments Checker v.5”，https://chsopensource.org/pigments-checker/。访问日期：2023-02-14。17. M. Müller、H. A. Rinia、M. Bonn、E. M. Vartiainen、M. Lisker 和 A. van Bel，“拥挤光谱区域中的定量多路复用相干反斯托克斯拉曼散射光谱学”，Multiphoton Microsc. Biomed. Sci. VII **6442**，21–29（2007）。18. E. M. Vartiainen、H. A. Rinia、M. Müller 和 M. Bonn，“从拥挤的相干反斯托克斯拉曼散射光谱中直接提取拉曼线型”，Opt. Express **14**，3622–3630（2006）。19. M. Mathlouthi 和 D. V. Luu，“水溶液中 D-果糖的激光拉曼光谱”，Carbohydr. Res. **78**，225–233（1980）。20. H. A. Rinia、M. Bonn 和 M. Müller，“拥挤光谱区域中的定量多路复用相干反斯托克斯拉曼散射光谱学”，The J. Phys. Chem. B **110**，4472–4479（2006）。

21. M. Müller 和 J. M. Schins，“利用多路复用相干反斯托克斯拉曼散射显微术成像脂质膜的热力学状态”，The J. Phys. Chem. B **106**, 3715–3723 (2002)。