Trace formulas for spatially averaged Raman scattering

Steven G. Johnson*

September 13, 2019

1 Overview

Incoherent Raman scattering can be modeled as a combination of two electromagnetic processes:

1. An incident field (or equivalent current source $\mathbf{J}_1$) produces an electric field $\mathbf{E}_1 e^{-i\omega_1 t}$ at a frequency ω_1. This solves the linear Maxwell equations $\mathcal{M}_1\mathbf{E}_1 = i\omega_1\mathbf{J}_1$ where $\mathcal{M}_1$ is the Maxwell (vector Helmholtz) operator $\mathcal{M}_1 = \nabla \times \mu_1^{-1}\nabla \times - \frac{\omega_1^2}{c^2}\varepsilon_1$, with $\varepsilon_1(\mathbf{x})$ and $\mu_1(\mathbf{x})$ being the electric and magnetic permeabilities at a frequency ω_1. (Usually $\mu_1 \approx 1$ at the frequencies of Raman processes.)

2. A molecule at position $\mathbf{x}_0$ with a Raman polarizability tensor α produces a dipole current density $\mathbf{J}_2 = \alpha\mathbf{E}_1(\mathbf{x}_0)\delta(\mathbf{x} - \mathbf{x}_0)e^{-i\omega_2 t}$ at a frequency ω_2, which produces an electric field $\mathbf{E}_2$ satisfying $\mathcal{M}_2\mathbf{E}_2 = i\omega_2\mathbf{J}_2$ where $\mathcal{M}_2$ is the Maxwell operator at the frequency ω_2.

Typically, we are interested in maximizing the power radiated into one or more directions/channels by $\mathbf{E}_2$ for a given $\mathbf{J}_1$. The difference $\omega_2 - \omega_1$ is the Raman shift, and usually $|\omega_2 - \omega_1| \ll |\omega_1|$.

On a computer, we discretize this problem (e.g. using finite elements) into a sequence of linear equations: $M_1 v_1 = b_1$, $b_2 = A v_1$, $M_2 v_2 = b_2$, and our objective is some Hermitian (probably positive-semidefinite) quadratic form $v_2^* Q v_2$ (i.e. $Q = Q^*$, where $*$ denotes the adjoint). We can then apply various shape-optimization techniques etcetera to optimize the spatial arrangement of dielectric/metallic materials (ε and μ).

The complication that these notes are intended to address is the situation where the Raman molecules are distributed randomly in some region (e.g. suspended in a fluid solution), which is typical in many experimental situations. In this case, what you are really interested in is the *average*

$$\langle v_2^* Q v_2 \rangle_\alpha = \left\langle b_1^* M_1^{-*} A^* M_2^{-*} Q M_2^{-1} A M_1^{-1} b_1 \right\rangle_\alpha , \tag{1}$$

*Notes following a discussion with Rasmus Christiansen.

where $\langle \cdots \rangle_\alpha$ denotes an average over all allowed positions $\mathbf{x}_0$ and orientations of the Raman molecule (possibly weighted by some nonuniform probability distribution); note that the only terms that depend on this are A and A^*. Naively, this could be computed by a multidimensional quadrature of Raman solves — that is, we repeat steps (1.)–(2.) for many different positions and orientations in order to explicitly average. However, this could be computationally expensive because of the many Maxwell solves that are required, and may be prohibitive in the context of shape optimization where the averaging must be repeated for many geometric shapes. Instead, we will adapt "trace formula" techniques that have been developed for similar problems in thermal radiation [1, 2] and spontaneous emission [3], where one must compute the average effect of many random current sources distributed throughout a volume. The basic trick (as reviewed in yet another related setting by [4]) is that $\langle v^* Q v \rangle = \mathrm{tr} \langle v^* Q v \rangle = \mathrm{tr}[Q \langle v v^* \rangle]$, and the correlation matrix $\langle v v^* \rangle$ usually simplifies in some dramatic way. In fact, the final expression (3) can be interpreted as a weighted Frobenius semi-norm of M_2^{-1}, which can potentially be computed by a variety of efficient iterative methods.

2 Raman trace formula

In the Raman case, we can use the trace trick to rearrange (1) in the form:

$$\langle v_2^* Q v_2 \rangle_\alpha = \mathrm{tr} \left[M_2^{-*} Q M_2^{-1} \left\langle A v_1 v_1^* A^* \right\rangle_\alpha \right]. \tag{2}$$

The goal now is to find a simple, tractable expression for $\left\langle A v_1 v_1^* A^* \right\rangle_\alpha$ (noting that $v_1 = M_1^{-1} b_1$ is a fixed vector independent of A).

Let us suppose that v_1 is simply an array of 3-component $\mathbf{E}_1(\mathbf{x})$ values at different points $\mathbf{x}$, in which case A for a given molecule location $\mathbf{x}_0$ is of the form:

$$A = \begin{pmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & \alpha & & & \\ & & & 0 & & \\ & & & & \ddots & \end{pmatrix},$$

and hence

$$A v_1 v_1^* A^* = \begin{pmatrix} \ddots & & & & & \\ & 0 & & & & \\ & & \alpha \mathbf{E}_0 \mathbf{E}_0^* \alpha^* & & & \\ & & & 0 & & \\ & & & & \ddots & \end{pmatrix}$$

where $\mathbf{E}_0 = \mathbf{E}_1(\mathbf{x}_0)$. We must first compute the average of the 3×3 matrix $\alpha \mathbf{E}_0 \mathbf{E}_0^* \alpha^*$ over all possible orientations of the molecule, i.e. the average of $Q \alpha Q^* \mathbf{E}_0 \mathbf{E}_0^* Q \alpha^* Q^*$ over all 3×3 rotations Q [sampled uniformly in SO(3)].

Call this matrix $\widehat{\alpha^2(\mathbf{x}_0)}$; we will return to the calculation of $\widehat{\alpha^2}$ below. Then we average over positions, resulting in:

$$\langle A v_1 v_1^* A^* \rangle_\alpha = D \underbrace{\begin{pmatrix} \widehat{\alpha^2(\mathbf{x}_1)} & & \\ & \widehat{\alpha^2(\mathbf{x}_2)} & \\ & & \ddots \end{pmatrix}}_{\widehat{A^2}} D^*$$

where the points $\mathbf{x}_i$ are the locations on our discretized grid/mesh, and $D = D^*$ is a diagonal real positive semidefinite matrix whose diagonal entries are the *square roots* of the probabilities of a molecule being at each point $\mathbf{x}_i$. (These probabiliies will be zero outside of the fluid where the Raman molecules are dissolved, for example.) Since $\widehat{\alpha^2}$ is necessarily Hermitian and positive semidefinite (because $\alpha \mathbf{E}_0 \mathbf{E}_0^* \alpha^*$ is), so is $\widehat{A^2}$, and hence it can be written as $\widehat{A^2} = \hat{A}\hat{A}^*$ for some factors $\hat{A}$ (e.g. the matrix square root or Cholesky factors) which are 3×3 block-diagonal like $\widehat{A^2}$. In fact, the factors $\hat{A}$ can be obtained analytically, as discussed below.[1]

Finally, we obtain

$$\boxed{\langle v_2^* Q v_2 \rangle_\alpha = \operatorname{tr}\left[\hat{A}^* D^* M_2^{-*} Q M_2^{-1} D \hat{A}\right]}. \tag{3}$$

which is manifestly the trace of a Hermitian matrix. If Q is semidefinite (usually a necessary consequence of passivity for power quadratic forms), then eq. (3) is equivalent to a weighted Frobenius seminorm $\|M_2^{-1}\|_W^2$ of M_2^{-1}, similar in spirit to [1, 2]. There are a variety of ways to estimate this trace (3), discussed below, since we can efficiently multiply any vector by the linear operator in the trace (the terms are either block-diagonal or correspond to Maxwell solves).

2.1 The rotation average $\widehat{\alpha^2}$

Let us now consider the computation of the average

$$\widehat{\alpha^2} = \langle Q \alpha Q^* \mathbf{E} \mathbf{E}^* Q \alpha^* Q^* \rangle_Q \,, \tag{4}$$

i.e. the average of $\alpha \mathbf{E} \mathbf{E}^* \alpha^*$ over all possible rotations $Q \in \mathrm{SO}(3)$ of α, sampled uniformly (assuming that the Raman molecules are equally likely to be in any orientation).

The simplest case is when α is isotropic: $\alpha = \alpha_0 I$ for some scalar α_0. In that case $\widehat{\alpha^2} = |\alpha_0|^2 \mathbf{E}\mathbf{E}^*$; since this is a rank-1 projection we can factorize it into non-square $\alpha_0 \mathbf{E}$ terms along the diagonal of $\hat{A}$.

[1] Although it is convenient to factorize $\widehat{A^2}$ like this in order to obtain a manifestly Hermitian matrix, it is not strictly necessary. One could instead leave it as $\widehat{A^2}$ and employ a trace formula related to a generalized eigenproblem on the support of $\operatorname{diag}(D)$.

For the general case of an anisotropic α, we could of course compute $\widehat{\alpha^2}$ by a brute-force numerical quadrature over rotations, and then perform e.g. a 3×3 Cholesky factorization to obtain the diagonal blocks of $\hat{A}$, but eq. (4) is simple enough to admit an analytical solution. In particular, Prof. Alan Edelman at MIT showed me how to obtain the result as a consequence of [5, Table IV]. The resulting matrix $\widehat{\alpha^2}$ has two eigenvalues $\alpha_\parallel^2|\mathbf{E}|^2$ and $\alpha_\perp^2|\mathbf{E}|^2$ for vectors parallel and perpendicular to $\mathbf{E}$, respectively, where:

$$\alpha_\parallel^2 = \frac{|\operatorname{tr}\alpha|^2 + \operatorname{tr}\left[\alpha^*\alpha + \bar{\alpha}\alpha\right]}{15},$$
$$\alpha_\perp^2 = \frac{\operatorname{tr}\left[\alpha^*\alpha\right]/3 - \alpha_\parallel^2}{2}.$$

(Note that if $\alpha = \alpha_0 I$, then $\alpha_\parallel^2 = |\alpha_0|^2$ and $\alpha_\perp^2 = 0$.) Therefore,

$$\widehat{\alpha^2} = \alpha_\parallel^2|\mathbf{E}|^2\frac{\mathbf{E}\mathbf{E}^*}{|\mathbf{E}|^2} + \alpha_\perp^2|\mathbf{E}|^2\left(I - \frac{\mathbf{E}\mathbf{E}^*}{|\mathbf{E}|^2}\right) = \left(\alpha_\parallel^2 - \alpha_\perp^2\right)\mathbf{E}\mathbf{E}^* + \alpha_\perp^2|\mathbf{E}|^2 I.$$

Because this matrix is generally full rank, we must factorize it into a product of square full-rank matrices, unlike the isotropic special-case above. For example, we can simply use the matrix square root

$$\hat{A} = \begin{pmatrix} \frac{\alpha_\parallel - \alpha_\perp}{|\mathbf{E}_1|}\mathbf{E}_1\mathbf{E}_1^* + \alpha_\perp|\mathbf{E}_1|I & & \\ & \frac{\alpha_\parallel - \alpha_\perp}{|\mathbf{E}_2|}\mathbf{E}_2\mathbf{E}_2^* + \alpha_\perp|\mathbf{E}_2|I & \\ & & \ddots \end{pmatrix},$$

where $\mathbf{E}_k = \mathbf{E}(\mathbf{x}_k)$.

3 Estimating the trace

Computationally, we require an iterative method to compute the trace (3), which is of the form $\operatorname{tr}[O^*QO]$ where

$$O = M_2^{-1}D\hat{A} \tag{5}$$

is the product of a Maxwell solve and two diagonal or block-diagonal matrices, and Q is the Hermitian (probably semidefinite) quadratic form for our power figure of merit (which should also be block diagonal, assuming a localized discretization basis, because power computation is local in space). There are a variety of ways to go about this.

The simplest trace-estimation method is a pure Monte-Carlo quadrature, also called Hutchinson's method and variations thereof [6]: $\operatorname{tr}[O^*QO] \approx \frac{1}{M}\sum_{i=1}^{M}(Oz_i)^*Q(Oz_i)$ for a random set of M vectors z_i (with several possible distributions [6]).

There are also approaches using a Krylov method, usually based on the Lanczos algorithm, in order to accelerate the Monte-Carlo process. In particular, the most famous algorithm is now known as Stochastic Lanczos Quadrature

(SLQ) [7]. As I understand it, SLQ is essentially Hutchinson's method for $\mathrm{tr}[f(X)]$ for some (SPD) matrix X, except that $z^* f(X) z$ is computed approximately using a Lanczos process for X—the goal is to avoid computing $f(X)z$ explicitly when f is expensive. In our case, the analogous acceleration would be to try to avoid explicitly computing $M_2^{-1} y$ for any vector y (assuming this requires an expensive iteration like GMRES): instead of applying an Krylov process to M_2 to compute $M_2^{-1} y$, we would use the Krylov basis to compute $(Oz)^* Q (Oz)$ directly. Since our Maxwell matrix M_2 is not Hermitian, we cannot use Lanczos, but there is a variant of SLQ that uses Golub–Kahan bidiagonalization which is applicable [7].

3.1 Rayleigh-quotient trace estimation

For the specific trace estimation arising in the Maxwell problem with optimized structures, M_2^{-1} may well be approximately low rank: the response is likely to be dominated by a small number of electromagnetic resonances [poles of $\mathcal{M}(\omega)^{-1}$ lying close to the real-ω axis]. In this case, we can simply apply an iterative method (e.g. Lanczos or LOPCG) to estimate a few of the largest eigenvalues of $O^* Q O$, in which case the trace is approximated by the sum of these eigenvalues. Moreover, for a Hermitian semidefinite matrix it turns out that there is a very convenient way to express the sum of the largest n eigenvalues by maximizing a block Rayleigh quotient [8]:

$$\mathrm{tr}[O^* Q O] \geq \max_{V \in \mathbb{C}^{m \times n}} \mathrm{tr}\left[(OV)^* Q (OV) (V^* V)^{-1} \right].$$

The "$\geq$" becomes equality for $n = m$, but for problems with strong resonances I expect that it will be a good approximation for $n \ll m$, perhaps even for $n = 1$ (for "small" structures where only one strong resonance "fits").

This Rayleigh-quotient maximization formula is especially attractive in the context of shape optimization, because it can be *combined with the shape optimization* itself. That is, instead of "nesting" the trace computation inside a larger shape optimization procedure, we can simply add V to the shape degrees of freedom and optimize over both V and the shape *simultaneously*. The full Raman shape-optimization problem becomes:

$$\boxed{\max_{\text{shape},\, V \in \mathbb{C}^{m \times n}} \mathrm{tr}\left[(OV)^* Q (OV) (V^* V)^{-1} \right]}, \tag{6}$$

where the shape affects M_1 (hence x_1 and $\hat{A}$), M_2, and D in O from (5). Maximizing this Rayleigh quotient has the nice property that, since you are maximizing a lower bound to the full trace, you are guaranteed to improve the full trace as well.

Note that the gradient of this trace expression with respect to V, in the CR-calculus sense, is simply $(I - VUV^*) O^* Q O V U$ where $U = (V^* V)^{-1}$ [8]. The gradient with respect to shape parameters can be obtained by the usual adjoint methods, since the $n \times n$ matrix that we are tracing is now computed explicitly.

References

[1] A. W. Rodriguez, M. T. H. Reid, and S. G. Johnson, "Fluctuating surface-current formulation of radiative heat transfer for arbitrary geometries," *Physical Review B*, vol. 86, p. 220302(R), December 2012.

[2] ——, "Fluctuating surface-current formulation of radiative heat transfer: Theory and applications," *Physical Review B*, vol. 88, p. 054305, August 2013.

[3] A. G. Polimeridis, M. T. H. Reid, W. Jin, S. G. Johnson, J. K. White, and A. W. Rodriguez, "Fluctuating volume-current formulation of electromagnetic fluctuations in inhomogeneous media: Incandescence and luminescence in arbitrary geometries," *Physical Review B*, vol. 92, p. 134202, October 2015.

[4] M. T. H. Reid, O. D. Miller, A. G. Polimeridis, A. W. Rodriguez, E. M. Tomlinson, and S. G. Johnson, "Photon torpedoes and Rytov pinwheels: Integral-equation modeling of non-equilibrium fluctuation-induced forces and torques on nanoparticles," *arXiv.org e-Print archive*, p. arXiv:1708.01985, 2017.

[5] R. Movassagh and A. Edelman, "Isotropic entanglement," *arXiv E-print Archive*, p. arXiv:1012.5039, 2010.

[6] H. Avron and S. Toledo, "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix," *Journal of the ACM*, vol. 58, no. 2, pp. 1–34, Apr. 2011. [Online]. Available: https://doi.org/10.1145/1944345.1944349

[7] S. Ubaru, J. Chen, and Y. Saad, "Fast estimation of $tr(f(a))$ via stochastic lanczos quadrature," *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 4, pp. 1075–1099, Jan. 2017. [Online]. Available: https://doi.org/10.1137/16m1104974

[8] S. Johnson and J. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Optics Express*, vol. 8, pp. 173–190, 2001. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=oe-8-3-173