

Mathematical and Physical Background for TSTC Lectures

Phillip L. Geissler

July 10, 2013

These brief notes are intended to refresh your memory on some basic concepts and tools pertinent to the upcoming lectures. Our presentations will make use of many of these results; given the short time available, we do not plan to discuss them on their own (though you should feel free to ask questions, and to point out inevitable typos).

Calculations labeled as exercises draw attention to manipulations you should be able to carry out, in most cases fairly readily. They should serve both as useful practice and as a guide to areas you might want to brush up on before we descend into the mines of theory next week.

1 Vector calculus

In three dimensions, the position of an object is specified by three coordinates, i.e., by a vector

$$\mathbf{r} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}}$$

Here, hats denote unit vectors in three orthogonal directions, and the scalar quantities x , y , and z indicate the corresponding components of \mathbf{r} . The length, or magnitude, of this vector is denoted $|\mathbf{r}|$ or when we are feeling lazy just r .

Scalar products, denoted by the symbol “ \cdot ”, should already be very familiar to you. For two generic vectors \mathbf{A} and \mathbf{B} ,

$$\mathbf{A} \cdot \mathbf{B} = A_x B_x + A_y B_y + A_z B_z,$$

where A_x denotes the component of \mathbf{A} along $\hat{\mathbf{x}}$, etc. Note that, as the name suggests, the result of this operation is a scalar quantity, i.e., just a number (not a vector or tensor).

There are other ways to multiply vectors. The “dyadic product” $\mathbf{D} = \mathbf{A} \mathbf{B}$ is a tensor quantity. For the things we will do to them, tensors are not scary beasts. In simple terms, a vector is a list of scalars; a tensor is just a list of vectors, which might be naturally written in matrix form. Components of \mathbf{D} are therefore specified by two indices (just like those of a simple matrix), e.g., $\mathbf{D}_{xz} = A_x B_z$.

A function $f(\mathbf{r})$ of position \mathbf{r} can vary in each of the three directions. There are thus many ways in which it can be differentiated (i.e., different partial derivatives). A compact way to describe the variation of $f(\mathbf{r})$ in space is through its gradient

$$\nabla f(\mathbf{r}) = \left(\frac{\partial f}{\partial x} \right)_{y,z} \hat{\mathbf{x}} + \left(\frac{\partial f}{\partial y} \right)_{x,z} \hat{\mathbf{y}} + \left(\frac{\partial f}{\partial z} \right)_{x,y} \hat{\mathbf{z}}.$$

Note that ∇f is a vector quantity, whose components specify the partial derivatives of f along the corresponding directions. You should be able to convince yourself that the net direction of ∇f is the direction in which $f(\mathbf{r})$ increases most rapidly. You should also be able to show that the rate of change of f along an arbitrary direction $\hat{\mathbf{A}}$ is $\hat{\mathbf{A}} \cdot \nabla f$. Finally, you should be able to make sense of quantities like $\nabla \cdot \nabla f \equiv \nabla^2 f$ (the Laplacian) and $\nabla \nabla f$ (a tensor – as practice you could work out its components).

2 Fourier transforms

A function $f(t)$ can be represented as a weighted sum of sines and cosines, provided it satisfies some modest conditions (which will not be belabored here). Using complex exponentials (recall that $e^{ia} = \cos a + i \sin a$),

$$f(t) = (2\pi)^{-1} \int_{-\infty}^{\infty} d\omega \hat{f}(\omega) e^{-i\omega t}$$

The integral here sums over different frequencies ω of oscillation, and $\hat{f}(\omega)$ indicates how strongly each frequency contributes to the signal $f(t)$. Orthogonality of oscillations at different frequencies allows us to obtain $\hat{f}(\omega)$ through a very similar expression:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} dt f(t) e^{i\omega t}$$

For later use, let's denote this linear operation on $f(t)$ as $\mathcal{F}[f(t)] = \hat{f}(\omega)$, the Fourier transform. We sometimes refer to $\hat{f}(\omega)$ as a representation of $f(t)$ “in reciprocal space”. Note that the relationship between $f(t)$ and $\hat{f}(\omega)$ is one-to-one, i.e., $f(t)$ has a unique representation in reciprocal space and $\hat{f}(\omega)$ determines a unique function of t .

Apology: We have already used hats for two different purposes, to denote unit vectors and to denote Fourier transforms. Context should make it clear which is meant in a given situation.

This way of representing a function brings many advantages. One is that differentiation, df/dt , becomes an algebraic operation in reciprocal space. To see this, just differentiate the Fourier representation

$$\frac{df}{dt} = (2\pi)^{-1} \int_{-\infty}^{\infty} d\omega \hat{f}(\omega) e^{-i\omega t} (-i\omega)$$

The uniqueness of Fourier transforms then tells us that $\mathcal{F}[df/dt] = -i\omega \hat{f}(\omega)$. This result of course implies that $\mathcal{F}[d^2 f/dt^2] = (-i\omega)^2 \hat{f}(\omega)$ and so on.

The variable t is suggestive of time as an independent variable, but we will also frequently transform functions of position \mathbf{r} . In more than one dimension, \mathbf{r} is a vector, requiring that we generalize the above results just a little:

$$f(\mathbf{r}) = (2\pi)^{-3} \int d\mathbf{k} \hat{f}(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{r}}$$

and

$$\hat{f}(\mathbf{k}) = \int d\mathbf{r} f(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}$$

Note that reciprocal space is also of more than one dimension. The “wavevector” \mathbf{k} thus indicates not just a wavelength ($\lambda = 2\pi/|\mathbf{k}|$) but also the direction $\hat{\mathbf{k}}$ along which the corresponding Fourier component oscillates.

As described in the previous section, there are many different ways to differentiate a function of a vector. As examples (that we will use), you should be able to show straightforwardly that

$$\mathcal{F}[\nabla f(\mathbf{r})] = -i\mathbf{k} \hat{f}(\mathbf{k}),$$

$$\mathcal{F}[\nabla^2 f(\mathbf{r})] = -k^2 \hat{f}(\mathbf{k}),$$

and

$$\mathcal{F}[\nabla \nabla f(\mathbf{r})] = -\mathbf{k} \mathbf{k} \hat{f}(\mathbf{k}).$$

3 Dirac's delta function

It is often convenient to have at one's disposal a function that is completely localized, i.e., one that vanishes except at a single point. Unfortunately, we often want that function also to have finite integrated strength, requiring that it be singular precisely at the point of interest. This singularity causes plenty of headaches, but it is worth the trouble.

The function in question is the Dirac delta function $\delta(x)$, defined by the two properties:

$$\delta(x) = 0 \quad \text{for } x \neq 0, \quad \text{and} \quad \int_{-\infty}^{\infty} dx \delta(x) = 1.$$

You should be able to convince yourself that $\delta(x)$ must diverge at $x = 0$ in order for the second equation to hold.

If you ever find yourself befuddled by operations involving delta functions, just replace them (in your mind, usually) by a function of finite width σ and unit integrated strength (e.g., a Gaussian distribution). Perform the operation in question with this nonsingular function, and then take the limit $\sigma \rightarrow 0$.

Integrals involving delta functions are usually very easy. For example, so long as $f(x)$ is nonsingular,

$$\int_{-\infty}^{\infty} dx \delta(x - x_0) f(x) = f(x_0).$$

You should be able to show that the Fourier transform of a delta function is very simple,

$$\mathcal{F}[\delta(t - t_0)] = e^{i\omega t_0}.$$

The symmetry and uniqueness of the Fourier transform then imply the useful identity

$$\int_{-\infty}^{\infty} d\omega e^{i\omega t} = 2\pi\delta(t)$$

If we wish to localize a function of the vector \mathbf{r} in all three directions, we'll need the generalization

$$\delta(\mathbf{r}) \equiv \delta(x)\delta(y)\delta(z)$$

Exercise: Show that

$$\nabla^2 \frac{1}{r} = -4\pi\delta(\mathbf{r})$$

Demonstrating that $\nabla^2 r^{-1} = 0$ for $\mathbf{r} \neq 0$ should be straightforward. Calculating its integral requires a little more finesse; the divergence theorem provides a short route.

Exercise: Show that

$$\mathcal{F}\left[\frac{1}{r}\right] = \frac{4\pi}{k^2}$$

(The result of the first exercise should come in handy.)

4 Statistics

Everyday life provides us with (sometimes misleading) intuition for randomness and probability. Most of us accept that we cannot anticipate the outcome of a single coin toss. But we expect that if we flip the same coin many, many times, the number of heads results *per flip* tends towards a consistent value (1/2 for an unbiased coin), with the typical deviation growing ever smaller as more flips are performed. It is this limiting frequency of an outcome that we call probability.

4.1 Probability distributions and averages

You should already be very familiar with basic properties of probability distributions, which characterize the frequencies of various outcomes for a particular measurement. As examples,

- A distribution $P(A)$ is properly normalized if summing over all possible outcomes A yields unity,

$$\sum_A P(A) = 1.$$

- The average, or expectation value, of a function of A is obtained by summing over appropriately weighted outcomes. We denote such an average with angled brackets:

$$\langle f(A) \rangle = \sum_A P(A) f(A).$$

In particular, the average of A itself ($f(A) = A$) is $\langle A \rangle = P(A_1)A_1 + P(A_2)A_2 + \dots$ (This expression only makes sense if the possible outcomes of A are numeric, which is assumed throughout these notes.)

- The probability of observing any one of a collection of outcomes $\{A_1, A_2, \dots, A_n\}$ is given by the sum of all individual probabilities:

$$\text{prob}(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

4.2 Correlations

A pair of measurements (say, of two properties A and B) may be correlated. This relationship is detailed by a joint probability $P(A, B)$ of observing a particular value of A *and* a particular value of B . If the two measurements are uncorrelated (i.e., the outcome A has no influence on the statistics of outcome B), then this joint distribution factorizes, $P(A, B) = P(A)P(B)$. The nature of such correlations can be quantified in several ways. A common approach is to compute the covariance, $\langle \delta A \delta B \rangle$, where $\delta A = A - \langle A \rangle$. (We will use this notation, in which δ indicates deviation of a random variable about its average, extensively). You should be able to show that this quantity vanishes, $\langle \delta A \delta B \rangle = 0$, if A and B are uncorrelated.

Conditional probability, detailing the statistics of one measurement A given that a particular outcome of B has been observed, provides another measure of correlation. We will denote this quantity as $P(A|B)$. It can be simply related to the joint and singlet probabilities:

$$P(A|B) = \frac{P(A, B)}{P(B)}.$$

Similarly,

$$P(B|A) = \frac{P(A, B)}{P(A)}.$$

If you have not worked with conditional probabilities before, these basic relationships deserve some rumination.

4.3 Moments and cumulants

The *moments* of a probability distribution $P(A)$ give information about its shape. The first moment, $\langle A \rangle$, indicates where $P(A)$ is centered. The second moment $\langle A^2 \rangle$ reflects the breadth of $P(A)$, but in a slightly indirect way: Note that if $P(A)$ is shifted to larger values of A , so that $\langle A \rangle$ increases, $\langle A^2 \rangle$ will also increase even if the distribution's width is unchanged. The same is true for higher moments ($\langle A^3 \rangle$, $\langle A^4 \rangle$, etc.) For this reason we often consider the *cumulants* of $P(A)$ instead.

The first cumulant, $C_1 = \langle A \rangle$, is the same as the first moment. The second cumulant, $C_2 = \langle (\delta A)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2$ (a.k.a. variance), directly indicates the width of $P(A)$; it is insensitive to shifting the distribution to the right or left. Similarly, the third cumulant $C_3 = \langle (\delta A)^3 \rangle$ directly characterizes how strongly the

distribution is skewed to one side or the other. Higher order cumulants have slightly more complicated forms, e.g., $C_4 = \langle (\delta A)^4 \rangle - 3\langle (\delta A)^2 \rangle^2$. They can be defined through a “cumulant generating function”,

$$Z(\lambda) = \langle e^{\lambda A} \rangle.$$

Specifically,

$$C_n = \left. \frac{d^n \ln Z}{d\lambda^n} \right|_{\lambda=0}.$$

It is good practice to confirm that the first four cumulants described above are consistent with this definition.

4.4 Continuous variables

Up to this point, we have implicitly been considering a random variable A with a discrete set of possible values. Many of the fluctuating quantities we will discuss are instead continuous degrees of freedom, e.g., a particle’s position, or the net dipole per molecule in a microscopic region in liquid water. Let us denote a generic continuous random variable as x , and its probability distribution as $p(x)$.

Most of the results described above generalize simply to this continuous case, replacing sums over discrete outcomes by integrals over x . The expectation value of a function $f(x)$, for instance, reads

$$\langle f(x) \rangle = \int dx p(x) f(x).$$

One substantive difference concerns the *meaning* of $p(x)$. In the discrete case, it is natural to ask a question like, “What is the probability that a coin lands heads up?” Its analog here, “What is the probability that x takes on a value x_0 ?”, is not a well posed question. If $p(x)$ is not singular, then the probability of observing *any* specific value of x is zero.

By $p(x)$, we actually mean something awkward like: “The probability of observing a value of x in the infinitesimal interval between x_0 and $x_0 + dx$ is $p(x_0)dx$.” In simpler terms, $p(x)$ is a probability *density*, i.e., a probability per unit x . Finite probabilities are obtained by integrating $p(x)$ over finite intervals.

I will nonetheless reserve the right to ask questions like, “What is the probability of observing a separation \mathbf{r} between two molecules?”, with the longer, awkward meaning implicit in the fact that \mathbf{r} is continuous. (Incidentally, the question “What is the free energy of binding between two molecules?” is ill-posed for the very same reason. That is why binding equilibrium constants have units, just as $p(x)$ has units of $1/x$.)

4.5 Gaussian distributions

The form of $p(x)$ most commonly encountered in physical chemistry is undoubtedly the Gaussian distribution, a.k.a., the bell-shaped curve:

$$p_G(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(x-x_0)^2/2\sigma^2},$$

where x_0 and σ are parameters defining the distribution’s center ($C_1 = \langle x \rangle = x_0$) and breadth ($C_2 = \langle \delta x^2 \rangle = \sigma^2$).

The reason for the ubiquitous appearance of Gaussian distributions is several-fold. Firstly, its mathematical tractability is nearly irresistible. For example, the cumulant generating function can be computed exactly:

$$\langle e^{\lambda x} \rangle = e^{\lambda x_0} e^{\lambda^2 \sigma^2 / 2}$$

Exercise: Show that this is true.

Exercise: Using this result, show that the cumulants of $p_G(x)$ all vanish except for the first two, i.e., $C_n = 0$ for $n > 2$.

The central limit theorem provides a more meaningful reason to become proficient with Gaussian distributions. This theorem concerns a quantity X which is a sum of N *independent* (i.e., uncorrelated) random variables:

$$X = x_1 + x_2 + x_3 + \dots + x_N$$

In the limit $N \rightarrow \infty$ the distribution of X tends towards a Gaussian form, with

$$\langle \delta X^2 \rangle = \sum_{i=1}^N \langle \delta x_i^2 \rangle.$$

This result holds *regardless* of how the x_i variables are distributed (with some modest restrictions on $p(x_i)$ that will not be belabored here). Whenever an observable has contributions from many independent sources, we should expect its statistics to be Gaussian to a first approximation.

5 Electrostatics

Forces between charged particles are centrally important to chemistry, appearing prominently in the electronic Schrödinger equation, in the theory of polar liquids and their solutions, in the determination of spectroscopic response functions, etc. The strength and direction of the fundamental Coulomb force should be very familiar to you. The force \mathbf{F}_1 on an ion with charge q_1 at position \mathbf{r}_1 , due to another charge q_2 at position \mathbf{r}_2 , acts along the direction of separation $\hat{\mathbf{r}}_{12} = (\mathbf{r}_1 - \mathbf{r}_2)/|\mathbf{r}_1 - \mathbf{r}_2|$,

$$\mathbf{F}_1 = k_C \frac{q_1 q_2}{r_{12}^2} \hat{\mathbf{r}}_{12},$$

with a strength that decays as the inverse square of the distance between them. There is much more to be said about electrostatics, but it all derives from this basic law.

A word about units: The proportionality constant k_C in Coulomb's law above has been left intentionally vague. Its value depends on how one bundles various fundamental constants with the charges and distances involved. In these notes, I will take the charges q to be given relative to $\sqrt{4\pi\epsilon_0}$, in which case $k_C = 1$.

Integrating the expression for \mathbf{F}_1 above, we can obtain the potential energy $u(r)$ associated with a pair of ions separated by a distance r :

$$u(r) = \frac{q_1 q_2}{r}.$$

5.1 Electric potentials and fields

A large collection of charged species can create very complicated electrostatic environments. Rather than detailing where each ion resides, we could characterize such an environment through the consequences of introducing a “test charge”.

Imagine inserting a point ion, with charge q , at position \mathbf{r} while holding all of the actual charges fixed. The energy change resulting from this insertion, per unit charge, defines the electric potential $\phi(\mathbf{r})$. Since Coulomb forces are pairwise additive, we can write the potential explicitly in terms of the positions \mathbf{r}_j and charges q_j of the actual particles (indexed by j):

$$\phi(\mathbf{r}) = \sum_j \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|}.$$

This potential can be written alternatively in terms of the charge density field $\rho(\mathbf{r}) \equiv \sum_j q_j \delta(\mathbf{r} - \mathbf{r}_j)$,

$$\phi(\mathbf{r}) = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}.$$

(Just plug in the definition of ρ and integrate to confirm.) Using the result of exercises above, you can now derive Poisson's equation,

$$\nabla^2 \phi(\mathbf{r}) = -4\pi \rho(\mathbf{r}).$$

If you have followed this calculation, you should recognize that Poisson's equation is just a fancy differential rewriting of Coulomb's law. It neatly hides the detailed specification of our collection of charges, but they are still lurking in $\rho(\mathbf{r})$.

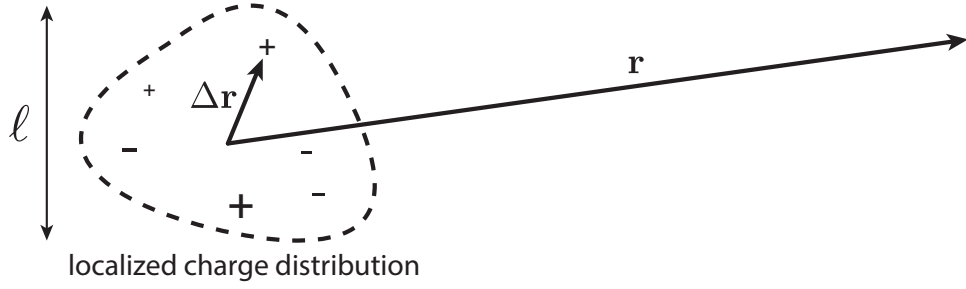


Figure 1: A charge distribution extending over a scale ℓ , whose effects are observed at a distant point \mathbf{r} .

The electric field \mathcal{E} generated by a collection of charges is similarly defined. Specifically, $\mathcal{E}(\mathbf{r})$ is the force, per unit charge, experienced by a point test ion inserted at \mathbf{r} . From the fundamental connection between energy and force, we know that the spatial variation of \mathcal{E} and ϕ contain the same information,

$$\mathcal{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})$$

5.2 Multipole expansion

Much as a probability distribution can be described in terms of its moments (or cumulants), a distribution of charges $\rho(\mathbf{r})$ can be equivalently described by a set of “multipole moments”. Such a description is particularly useful when probing the influence of $\rho(\mathbf{r})$ from a large distance away.

Imagine that a charge distribution is localized to a volume of scale ℓ near the origin. (See Fig. 1.) For example, we might have in mind a molecule of size ℓ (conceived as a distribution of nuclear and electronic charge) or a handful of nearby molecules within a distance ℓ of each other.

Let us compute the electric potential ϕ generated by ρ at a distant point \mathbf{r} , with $r \gg \ell$. We showed above that

$$\phi(\mathbf{r}) = \int d\Delta\mathbf{r} \frac{\rho(\Delta\mathbf{r})}{|\mathbf{r} - \Delta\mathbf{r}|},$$

where we have renamed the integration variable to emphasize that $\rho(\Delta\mathbf{r})$ is only nonzero for points $\Delta\mathbf{r}$ that are (comparably) very close to the origin. Since Δr is much smaller than r , it is tempting to introduce the Taylor expansion about $\Delta\mathbf{r} = 0$,

$$\frac{1}{|\mathbf{r} - \Delta\mathbf{r}|} = \frac{1}{r} - \Delta\mathbf{r} \cdot \nabla \frac{1}{r} + \frac{1}{2}(\Delta\mathbf{r} \cdot \nabla)^2 \frac{1}{r} + \dots$$

Combining these two expressions, we can write $\phi(\mathbf{r})$ as

$$\phi(\mathbf{r}) = \frac{Q}{r} - \mathbf{m} \cdot \nabla \frac{1}{r} + \dots,$$

where $Q = \int d\Delta\mathbf{r} \rho(\Delta\mathbf{r})$ is the net charge of ρ , and $\mathbf{m} = \int d\Delta\mathbf{r} \rho(\Delta\mathbf{r}) \Delta\mathbf{r}$ is its dipole, etc. This is a lovely little expression. For sufficiently large r , the series is strongly dominated by the lowest order terms. From a distance, the most easily resolved aspect of a localized charge distribution is thus its net charge Q . If you squint a little more closely, you might notice that it also bears a net dipole \mathbf{m} . Looking in still greater detail, you could detect higher order multipoles, but they are generally important only if lower order contributions vanish (e.g., if $Q = 0$ and $\mathbf{m} = 0$, we should probably pay attention to the quadrupole). This caricature of ρ is much less useful if we view the charge distribution from up close, in which case the series might converge very slowly.

Exercise: Work out the next couple of terms in this multipole expansion of ϕ , which involve the quadrupole and octupole moments.

From the expansion, we can also read off the potential generated by a point dipole at position \mathbf{r}_0 as $\phi_{\text{dipole}}(\mathbf{r}) = -\mathbf{m} \cdot \nabla |\mathbf{r} - \mathbf{r}_0|^{-1}$. The corresponding electric field is $\mathcal{E}_{\text{dipole}}(\mathbf{r}) = -\mathbf{m} \cdot \nabla \nabla |\mathbf{r} - \mathbf{r}_0|^{-1}$.

Let's ask a slightly different question about our localized charge distribution. Imagine that we subject it to an external potential ϕ_{ext} generated by some set of distant charges. What is the resulting energy U of interaction between ρ and ϕ_{ext} ?

From the definition of electric potential, we can begin with

$$U = \sum_j q_j \phi_{\text{ext}}(\mathbf{r}_j) = \int d\Delta\mathbf{r} \rho(\Delta\mathbf{r}) \phi_{\text{ext}}(\Delta\mathbf{r}).$$

Since $\phi_{\text{ext}}(\mathbf{r})$ should not vary much over the scale ℓ (why?), it is natural here to Taylor expand the external potential,

$$\phi_{\text{ext}}(\Delta\mathbf{r}) = \phi_{\text{ext}}(0) + \Delta\mathbf{r} \cdot \nabla \phi_{\text{ext}}(0) + \dots$$

Inserting this expansion, we obtain

$$U = Q\phi_{\text{ext}}(0) - \mathbf{m} \cdot \mathcal{E}_{\text{ext}}(0) + \dots$$

(You should be able to quickly confirm this result.) We can again read off an important property of point dipoles, namely that they like to align with external electric fields.

5.3 Dipole-dipole interactions

Interactions between pairs of dipoles will play an important role in developing a theory for polar liquids. We now have the ingredients to quickly write down the energy of this interaction.

In particular, we have shown that a dipole \mathbf{m}_1 at position \mathbf{r}_1 generates an electric field

$$\mathcal{E}(\mathbf{r}) = \mathbf{m}_1 \cdot \nabla \nabla \frac{1}{|\mathbf{r} - \mathbf{r}_1|}$$

at position \mathbf{r} . We have also shown that a dipole \mathbf{m}_2 at position \mathbf{r}_2 interacts with an external field $\mathcal{E}(\mathbf{r})$ with energy $u = -\mathbf{m}_2 \cdot \mathcal{E}$. Combining these results, we have

$$u_{\text{dip-dip}} = -\mathbf{m}_1 \cdot \nabla_2 \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \cdot \mathbf{m}_2 = \mathbf{m}_1 \cdot \nabla_1 \nabla_2 \frac{1}{|\mathbf{r}_2 - \mathbf{r}_1|} \cdot \mathbf{m}_2,$$

where ∇_1 and ∇_2 indicate differentiation with respect to \mathbf{r}_1 and \mathbf{r}_2 , respectively. Defining a dipole interaction tensor

$$\mathbf{T}(\mathbf{r}, \mathbf{r}') \equiv \nabla \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|},$$

we can write the interaction energy more compactly as $u_{\text{dip-dip}} = \mathbf{m}_1 \cdot \mathbf{T}(\mathbf{r}_1, \mathbf{r}_2) \cdot \mathbf{m}_2$.

Exercise: Show that this definition of $\mathbf{T}(\mathbf{r}_1, \mathbf{r}_2)$ is consistent with the somewhat more common expression

$$\mathbf{T}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{r_{12}^3} (\mathbf{I} - 3 \hat{\mathbf{r}}_{12} \hat{\mathbf{r}}_{12}),$$

where \mathbf{I} is the identity tensor.