Computational Science WS 16/17

November 11, 2016

Date	Lectures (Frank Noé)	Date	Lectures (Rupert Klein)
19. Oct	Overview 1	9. Dez	
21. Oct	Overview 2	14. Dez	
26. Oct	Overview 3	16. Dez	
28. Oct	Electrodynamics (JHP)	4. Jan	
2. Nov	Electrostatics and Gravity (JHP)	6. Jan	
4. Nov	Particle-particle methods, cutoff, short/long ranged	11. Jan	
9. Nov	Efficient neighbor search and implementation (Moritz)	13. Jan	
11. Nov	Poisson equation, uniqueness, solution, relaxation methods	18. Jan	
16. Nov	FFT	20. Jan	
18. Nov	Ewald	25. Jan	
23. Nov	Numerics and PME	27. Jan	
25. Nov	Metropolis-Hastings Monte-Carlo	1. Feb	
30. Nov	PCA	3. Feb	
2. Dez	Multipole method I (JHP)	8. Feb	
7. Dez	Multipole method II (JHP)	10. Feb	
13.-17. Feb: Exams			

1 Introduction

1.1 Examples of N-body systems

(slides on molecular dynamics and gravitational dynamics)

1.2 Basic notation

Consider a N-body system in which each body is represented by a point particle with position

$$\mathbf{r}_i = (r_{ix},\, r_{iy},\, r_{iz})^\top$$

Here we will generally consider "real-world" systems where the particles are located in three-dimensional Euclidean space, $\mathbf{r}_i \in \mathbb{R}^3$. Nothing prevents us from considering other spaces, such as two-dimensional spaces. Most equations will not change qualitatively when working in different dimensions, only a factors that depend on the dimensionality will be different. If the system is dynamic, these positions also depend on time, i.e. $\mathbf{r}_i(t) = (r_{ix}(t),\, r_{iy}(t),\, r_{iz}(t))^\top$, but we will consider that in the next section. For simplicity of notation, we write all positions of the N-body system in a single $3N$-dimensional vector

$$\mathbf{r} = (\mathbf{r}_1,\, ...,\, \mathbf{r}_N)^\top \in \Omega_r.$$

Here, Ω_r is the *configuration space* or *position space* of the system and its elements $\mathbf{r}$ are the possible *configurations* of the system. If our system is dynamic we can furthermore assign a velocity to each particle

$$\mathbf{v}_i = (v_{ix},\, v_{iy},\, v_{iz})^\top$$

and write the entire velocity vector of the system as

$$\mathbf{v} = (\mathbf{v}_1,\, ...,\, \mathbf{v}_N)^\top.$$

In dynamical systems each particle can be assigned a mass m_i. It is often more convenient to use the momentum instead of the velocity. The momentum is defined by

$$\mathbf{p}_i = m_i\mathbf{v}_i = (m_i v_{ix},\, m_i v_{iy},\, m_i v_{iz})^\top \in \Omega_p.$$

where Ω_p is the *momentum space* which contains all possible momentum vectors of the system. It is sometimes convenient to define the mass matrix

$$\mathbf{M} = \mathrm{diag}(m_1,\, m_1,\, m_1,\, m_2,\, m_2,\, m_2,\, ...,\, m_N,\, m_N,\, m_N)$$

with which we can write

$$\mathbf{p} = \mathbf{M}\mathbf{v}.$$

1.3 Energy and Force

Forces are fundamental aspects of physical systems. Let

$$\mathbf{f}(\mathbf{r}) = (f_x(\mathbf{r}),\, f_y(\mathbf{r}),\, f_z(\mathbf{r}))$$

be the force acting on a particle at position $\mathbf{r}$. Here we will be interested especially in *conservative forces*. A force is conservative when it meets any of these three equivalent conditions:

1. The force can be written as the negative gradient of a **potential**, Φ

   $$\mathbf{f}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$$

   where Φ is a scalar.

2. The curl of $\mathbf{f}$ is the zero vector:

   $$\nabla \times \mathbf{f} = \mathbf{0}$$

3. The net work done by the force when moving a particle in a cyclical path is zero:

$$W = \oint \langle \mathbf{f}, \, \mathrm{d}\mathbf{r} \rangle = 0$$

In classical mechanics, if we treat all masses explicitly, forces are conservative. Nonconservative forces can arise by considering only a subset of the masses and treating the other ones implicitly, i.e. by some statistical method. Friction and non-elastic material stress, and general relativity are nonconservative.

Here we are especially interested in the first formulation: The force can be written as the negative gradient of a potential Φ. This is a very powerful concept: The force is a relatively complex object: $\mathbf{f}(\mathbf{r})$ is a vector field, i.e. it is defined by a vector (in our case in $\mathbb{R}^3$) at every configuration $\mathbf{r}$. For conservative forces, it is sufficient to know a scalar field Φ, i.e. a scalar at every configuration $\mathbf{r}$, and we can derive the force field from it by taking the gradient at every configuration needed.

Fortunately it gets even simpler, because for many physical systems, Φ has a relatively simple mathematical form, or at least we have a relatively simple mathematical model which approximates reality well.

Example 1: Gravity. Consider a system of 2 masses, e.g. earth and moon. The ideal gravitational force of point mass m_2 at position $\mathbf{r}_2$ due to point mass m_1 at position $\mathbf{r}_1$ is given by:

$$\mathbf{f}_2(\mathbf{r}) = -G\frac{m_1 m_2}{|\mathbf{r}_{12}|^2}\hat{\mathbf{r}}_{12}$$

where $G = 6.674 \times 10^{-11}(\mathrm{Nm/kg})^2$ is the gravitational constant,

$$\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$$

is the vector pointing from object 1 to object 2, and the corresponding normal vector is

$$\hat{\mathbf{r}}_{12} = \frac{\mathbf{r}_{12}}{|\mathbf{r}_{12}|}.$$

Obviously, the component at mass 1 is

$$\mathbf{f}_1(\mathbf{r}) = -G\frac{m_1 m_2}{|\mathbf{r}_{21}|^2}\hat{\mathbf{r}}_{21},$$

i.e. of equal magnitude but pointing in the opposite direction. This is also necessary because forces must be balanced. If the bodies are close relative to their diameters, then we cannot treat them as point masses, but the above force must be obtained an an integral over all mass elements of the two bodies. Gravitational forces are conservative, and the gravitational potential for two point masses is

$$\Phi(\mathbf{r}) = G\frac{m_1 m_2}{|\mathbf{r}_{12}|},$$

such that $\mathbf{f}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$ is fulfilled. Note that we have the boundary condition $\lim_{|\mathbf{r}_{21}|\to\infty}\Phi(\mathbf{r}) = 0$, such that the integration constant is zero. The force felt by mass i within an N-body system of masses is given by:

$$\mathbf{f}_i(\mathbf{r}) = -m_i G\sum_{j\neq i}\frac{m_j}{|\mathbf{r}_{ji}|^2}\hat{\mathbf{r}}_{ji},$$

and the total potential energy of the system is given by

$$\Phi(\mathbf{r}) = G\sum_{i=1}^{N-1}\sum_{i=j+1}^{N}\frac{m_i m_j}{|\mathbf{r}_{ij}|},$$

While the gravitational force is the weakest of all forces, it is very long ranged. The interaction energy decays as $|\mathbf{r}_{12}|$, i.e. it is still active when strong short-ranged forces, such as the nuclear forces, have decayed.

Example 2: Molecular dynamics. In molecular systems, the gravitational force is so minor that it can be neglected. At these lengthscales electrodynamic forces play the main role. The simplest type

of interaction is the Coulomb interaction between charged particles, such as ions. Two point charges q_1 and q_2 at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ have the interaction potential

$$\Phi_{\text{Coulomb}}(\mathbf{r}) = \frac{q_1 q_2}{k_e \left|\mathbf{r}_{12}\right|}$$

where

$$k_e = \frac{1}{4\pi\varepsilon_0} = \frac{c_0^2 \mu_0}{4\pi} = c_0^2 \times 10^{-7}\ \mathrm{H \cdot m^{-1}} = 8.987 \times 10^9\ \mathrm{Nm^2 C^{-2}}$$

is Coulomb's constant. The above expression is using the Maxwell relation and contains other constants, such as the speed of light c_0, the magnetic permeability of vacuum μ_0, and electric permittivity of vacuum ϵ_0. So the Coulomb energy is unfavorable (positive) if charges have equal sign and favorable (negative) if they have opposite sign. Like the gravitational energy, it goes as $1/\left|\mathbf{r}_{12}\right|$, so it is long ranged, but its magnitude is many orders of magnitude larger than the graviational force for molecular systems, such that we usually neglect the gravitational force at this scale.

If charges are opposite, then the energy goes to $-\infty$ when the distance goes to 0. That is obviously a problem, because in reality two atoms cannot just collapse. When molecules are close, there are other forces that are also relevant. The attractive van der Waals force, or dispersion force, and the repulsive Pauli exclusion. Often, the energy of these components is qualitatively modeled by the Lennard-Jones potential:

$$\Phi_{\text{LJ}}(\mathbf{r}) = 4\epsilon_{12}\left[\left(\frac{\sigma_{12}}{|\mathbf{r}_{12}|}\right)^{12} - \left(\frac{\sigma_{12}}{|\mathbf{r}_{12}|}\right)^{6}\right]$$

where ϵ_{12} is the well depth at the minimum and σ_{12} is the effective atom radius. The total potential of N atoms using both ingredients is then given by

$$\Phi(\mathbf{r}) = \sum_{\text{pairs } i,j} \frac{q_i q_j}{k_e \left|\mathbf{r}_{ij}\right|} + \sum_{\text{pairs } i,j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{|\mathbf{r}_{ij}|}\right)^{12} - \left(\frac{\sigma_{ij}}{|\mathbf{r}_{ij}|}\right)^{6}\right].$$

Note that molecular dynamics models use more terms, for example they have terms that affect bond lengths, and angles between bonds and torsions.

In a classical molecular dynamics simulation, the conservative force is given by the gradient $\mathbf{f}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$.

Kinetic energy: While the potential energy of the system is often based on a model, the kinetic energy of the system is well defined for an N-body system. The kinetic energy of each single particle is given by

$$K(\mathbf{v}_i) = \frac{1}{2} m_i \left|\mathbf{v}_i\right|^2 = \frac{1}{2m_i}\left|\mathbf{p}_i\right|^2 = K(\mathbf{p}_i)$$

For the entire system we have

$$K(\mathbf{p}) = \frac{1}{2}\sum_i \frac{1}{m_i}\left|\mathbf{p}_i\right|^2 = \frac{1}{2}\mathbf{p}^\top \mathbf{M}^{-1}\mathbf{p}$$

where we have used the definition of the mass matrix $\mathbf{M}$ introduced above.

1.4 Dynamics

Next we turn to dynamics, i.e. we are interested in how the N-body system - be it a graviational system such as a Galaxy, or a molecular system such a protein - evolves in time t. We have silently assumed that the state variables depend on time, e.g. $\mathbf{q}(t)$, $\mathbf{p}(t)$. Now we will make this time dependence explicit. The basic equation needed for the dynamics of N-body system is Newtons second law, which in scalar variables is $f = ma$, where f is the force, m is the mass and $a = dv/dt = d^2x/dt^2$ is the acceleration that is the time derivative of the velocity and the second time derivative of the position. For an N-body system we employ the vector notation:

$$\mathbf{f} = \mathbf{M}\mathbf{a}$$

with the acceleration

$$\mathbf{a} = \frac{d}{dt}\mathbf{v} = \mathbf{a} = \mathbf{M}^{-1}\frac{d}{dt}\mathbf{p}$$

we get just

$$\mathbf{f} = \frac{d}{dt}\mathbf{p}$$

and using

$$\mathbf{p} = \mathbf{M}\frac{d}{dt}\mathbf{r}$$

we have

$$\mathbf{f} = \mathbf{M}\frac{d^2}{dt^2}\mathbf{r}.$$

Thus we can write:

$$\frac{d^2}{dt}\mathbf{r}_i = \frac{\mathbf{f}_i}{m_i}$$

Integrating the equation of motion over t yields the solution of the dynamical trajectory:

$$\mathbf{p}_i(t) = \mathbf{p}_i(0) + \int_{t'=0}^{t} \mathbf{f}_i(t')\, dt' \tag{1}$$

$$\mathbf{r}_i(t) = \mathbf{r}_i(0) + \mathbf{M}^{-1}\int_{t'=0}^{t} \mathbf{p}_i(t')\, dt' \tag{2}$$

Thus, given a starting condition $\mathbf{r}_i(0)$, $\mathbf{p}_i(0)$, these equations provide the entire dynamical trajectory $\mathbf{q}_i(t)$, $\mathbf{p}_i(t)$ for any time $t \in 0^+$. Since, for Newtonian dynamics we have time-reversibility:

$$\mathbf{r}_i(0),\, \mathbf{p}_i(0) \to \mathbf{r}_i(t),\, \mathbf{p}_i(t)$$
$$\mathbf{r}_i(t),\, -\mathbf{p}_i(t) \to \mathbf{r}_i(0),\, \mathbf{p}_i(0)$$

where $\to$ indicates time evolution, we can also go back to any past time $t \in 0^-$ by inverting the momenta and integrating normally.

Unfortunately, Eqs. (1-1) can only be solved analytically in very few cases. For example, try solving them for an harmonic oscillator, i.e. using the potential $\Phi(x) = \frac{1}{2}kx^2$. In general, the integrals are not be analytically tractable for N-body systems, so we need to employ a numerical integration scheme, in short called **integrator**.

The simplest choice is the **explicit Euler**, which we will only show for pedagogical reasons. Consider approximating the integral with $S = t/\Delta t$ little steps of length Δt:

$$\begin{aligned}\mathbf{p}(t) &\approx \mathbf{p}(0) + \sum_{n=0}^{S-1} \Delta t\, \mathbf{p}(n) \\ &= \mathbf{p}(0) + \sum_{n=0}^{S-2} \Delta t\, \mathbf{f}(n) + \Delta t\, \mathbf{f}(S-1) \\ &= \mathbf{p}(t-\Delta t) + \Delta t\, \mathbf{f}(t-\Delta t).\end{aligned}$$

And likewise we can approximate

$$\mathbf{r}_i(t) = \mathbf{r}_i(t-\Delta t) + \Delta t\, \mathbf{M}^{-1}\mathbf{p}(t-\Delta t).$$

WARNING: The explicit Euler scheme is really bad for our purposes (see computer-oriented Mathematics II). It will quickly diverge from the true solution, and worse than that it will be unstable (i.e. lead to a exponential increase of the total energy in the system).

A more stable and energy-conserving integrator is the **velocity Verlet** method, which works as follows:

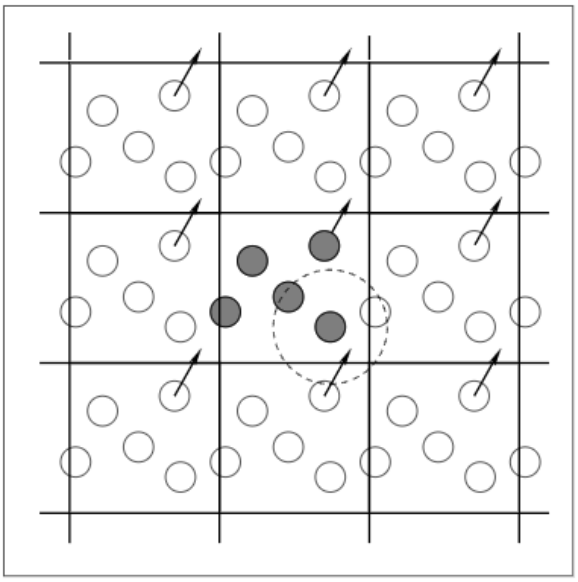
$$\mathbf{p}\left(t + \frac{\Delta t}{2}\right) = \mathbf{p}(t) + \frac{\Delta t}{2}\,\mathbf{f}(t).$$

$$\mathbf{r}\left(t + \frac{\Delta t}{2}\right) = \mathbf{r}(t) + \Delta t \mathbf{M}^{-1}\mathbf{p}\left(t + \frac{\Delta t}{2}\right).$$

$$\mathbf{p}\left(t + \Delta t\right) = \mathbf{p}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2}\,\mathbf{f}(t + \Delta t).$$

1.5 Boundary conditions

Periodic boundary conditions:



(Figure taken from http://wwwabi.snv.jussieu.fr/jompo/Public/PSF/IntroductionMolecularDynamics.pdf)

1.6 Stochastic dynamics and statistical mechanics

The classical mechanics using Newton's equation of motion described above is the mechanics of a **closed system**. A closed system is characterized by the **conservation of total energy in time**, i.e.:

$$H(\mathbf{r},\, \mathbf{p}) = \Phi(\mathbf{r}) + K(\mathbf{p})$$

is constant - the fraction of potential and kinetic energy changes over time, but the total energy doesn't. Such as system is also said to be in the **microcanonical ensemble**. A microscopical ensemble, as it is closed, is also deterministic - given an initial setting for $(\mathbf{r}(0),\, \mathbf{p}(0))$, the entire trajectory in the future and past is determined.

Especially microscopic systems, such as molecular systems, are usually not closed, but rather exchange energy (e.g. heat or particles) with their surrounding medium. In such systems, we are thus usually interested in other ensembles. The simplest case is the **canonical ensemble**, which is characterized by constant volume, temperature and number of particles. Imagine a container of fixed size, with boundaries that prevent particles from entering or leaving. However that container is kept at a constant temperature, and to do so it must exchange heat with the environment. The **temperature** can be defined *via* the mean kinetic energy K/N, by:

$$T = \frac{2}{3k_B}\frac{K}{N}$$

k_B is the Boltzmann constant. Note that the temperature is an ensemble property, i.e. the number of particles N must be very large in order to speak about temperature in a meaningful way.

From a statistical viewpoint, you can relate the microcanonical and the canonical ensemble in the following way. Imagine a distribution of microcanonical systems, each with a constant total energy $H(\mathbf{r},\, \mathbf{q})$. The correct distribution turns out to be the Boltzmann distribution, i.e. for each pair $(\mathbf{r},\, \mathbf{q})$ we have the following probability of occuring in the ensemble:

$$\mathbb{P}(\mathbf{r},\, \mathbf{q}) \propto \exp\left(-\frac{H(\mathbf{r},\, \mathbf{q})}{k_B T}\right).$$

This is the reason for the name microcanonical, as it could be considered as a part of the canonical ensemble.

However, this is a purely statistical viewpoint in practice we usually can't afford to simulate a large set of microcanonical simulations and weigh them. Instead we run only one simulation, but instead of keeping the energy constant, we couple the simulation to a **thermostat**, which has the purpose of allowing $H(\mathbf{r}, \mathbf{q})$ to fluctuate over time, in such a way that $\mathbb{P}(\mathbf{r}, \mathbf{q})$ is sampled from the Boltzmann distribution. Since we cannot simulate an infinite number of particles, we can in practice not keep the temperature T exactly constant, but it will rather also fluctuate around a fixed mean temperature. However the main property is to ensure that $\mathbb{P}(\mathbf{r}, \mathbf{q})$ is sampled correctly.

Stochastic Dynamics: There are many different thermostats and other coupling algorithms (barostates, chemostats, etc.). Here we will briefly discuss two commonly used integration schemes that include a thermostat, i.e. these are full algorithms to simulate molecular dynamics in the NVT ensemble. The first scheme is multidimensional Langevin dynamics

$$\mathbf{M}\frac{d^2}{dt^2}\mathbf{r} = \mathbf{f}(\mathbf{r}) - \Gamma(\mathbf{r})\frac{d}{dt}\mathbf{r} + \mathcal{D}(\mathbf{r})\boldsymbol{\eta}(t)$$

where Γ is the matrix of Stokes friction coefficients and $\mathcal{D}$ is the matrix of noise intensities. $\boldsymbol{\eta}(t)$ is a vector of uncorrelated Gaussian random numbers. Note the first part of the equation is equal to Newtonian dynamics, and the function of the last two terms is to act as a thermostat. While the above form is quite general, we can in practice often choose a relatively simply form for the friction and noise coefficients, namely:

$$\mathbf{M}\frac{d^2}{dt^2}\mathbf{r} = \mathbf{f}(\mathbf{r}) - \gamma\mathbf{M}\frac{d}{dt}\mathbf{r} + \sqrt{2k_BT\gamma}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\eta}(t)$$

where γ is an isotropic friction constant. If we simulate all particles that are physically meaningful for our system - such as explicitly-solvated molecular dynamcis, we choose γ relatively small, because the particles should not feel "real" friction in addition to the particles they actually collide with, so the whole purpose of γ is to couple our system to the heat bath, and we want to avoid slowing down our system artificially. Note that for $\gamma = 0$, we recover Newtonian dynamics.

In the overdamped limit, the equation can be written as the overdamped Langevin, or Smoluchowski, or Brownian dynamics equation:

$$\frac{d}{dt}\mathbf{r} = \gamma^{-1}\mathbf{M}^{-1}\mathbf{f}(\mathbf{r}) + \sqrt{\frac{2k_BT}{\gamma}}\mathbf{M}^{-\frac{1}{2}}\boldsymbol{\eta}(t)$$

Note that in the overdamped limit the motion is no longer intertial, which is why the accelerations do not appear anymore. Under certain circumstances we can use the so-called fluctuation dissipation theorem and write

$$\frac{d}{dt}\mathbf{r} = \frac{1}{k_BT}\mathbf{D}\mathbf{f}(\mathbf{r}) + \sqrt{2}\mathbf{D}^{\frac{1}{2}}\boldsymbol{\eta}(t)$$

where $\mathbf{D} = \mathrm{diag}(d_{1x}, d_{1y}, d_{1z}, ..., d_{Nx}, d_{Ny}, d_{Nz})$ is a diagonal matrix of diffusion coefficients. Note that if we have dynamics of rigid bodies that are not just point particles, $\mathbf{D}$ will not be diagonal, but rather be replaced by a diffusion tensor.

1.7 Monte Carlo simulation

Let us reconsider the Boltzmann distribution, and insert the definitions of potential and kinetic energy:

$$\begin{aligned}\mathbb{P}(\mathbf{r}, \mathbf{q}) &\propto \exp\left(-\frac{\Phi(\mathbf{r})}{k_BT} - \frac{K(\mathbf{p})}{k_BT}\right). \\ &\propto \exp\left(-\frac{\Phi(\mathbf{r})}{k_BT}\right)\exp\left(-\frac{K(\mathbf{p})}{k_BT}\right).\end{aligned}$$

Now we're not interested in the statistics of momenta anymore, so we integrate them out:

$$\mathbb{P}(\mathbf{r}) \propto \exp\left(-\frac{\Phi(\mathbf{r})}{k_BT}\right).$$

The integral over the momentum term results in a constant (the partition sum of momenta), but it can be ignored for our purpose. In short we just average over momenta for each position $\mathbf{r}$. Now we are interested in sampling this probability density. A conceptually simple algorithm is Monte Carlo simulation. Here we show a very simple version of Metropolis Monte Carlo. We proceed in the following steps:

1. Start with initial configuration $\mathbf{r}^{(0)}$ and set $k = 0$.
2. For $k = 1$, ..., K:
   (a) Sample random vector $\boldsymbol{\eta} \in \mathbb{R}^N$ with $\eta_i \sim \mathcal{N}(0, \sigma^2)$, i.e. an isotropic multivariate Gaussian distribution with variances σ^2.
   (b) Propose new configuration $\mathbf{r}' = \mathbf{r}^{(k)} + \boldsymbol{\eta}$.
   (c) Accept new configuration with probability

$$p_{\text{acc}} = \min\left\{1, \exp\left(-\frac{\Phi(\mathbf{r}^{'}) - \Phi(\mathbf{r}^{(k)})}{k_B T}\right)\right\}$$

   If accepted, set $\mathbf{r}^{(k+1)} = \mathbf{r}^{'}$, otherwise $\mathbf{r}^{(k+1)} = \mathbf{r}^{(k)}$.

This algorithm will in the long run sample from the correct probability density $\mathbb{P}(\mathbf{r})$. The algorithm can be modified in many different ways. In particular, we have a lot of freedom in designing Monte-Carlo moves (2a-b). Instead of adding random vectors from a symmetric distribution, we could propose very different steps, and as long as we can evaluate the probability of proposing a given step, we can always account for that in the acceptance step (2c). This more general variant is called Metropolis-Hastings Monte Carlo. We will not go deeper into this subject right now, but just mention that Monte Carlo methods are used ubiquitously in the computational sciences, and it is extremely useful to know this technology.

1.8 Optimization

Now we shall be interested in the question which system coordinates are most probable, i.e. we want to know what are the configurations $\mathbf{r}$ for which $\mathbb{P}(\mathbf{r})$ is maximum? So instead of running dynamics, we could use the following algorithm:

1. Start with initial configuration $\mathbf{r}^{(0)}$ and set $k = 0$.
2. Increment $k = 1$, ..., K until convergence criterion met:
   (a) Move configuration $\mathbf{r}^{(k)} \to \mathbf{r}^{(k+1)}$ in such a way that $\Phi(\mathbf{r}^{(k+1)}) \leq \Phi(\mathbf{r}^{(k)})$.
3. Return $\mathbf{r}^{(K)}$, $\Phi(\mathbf{r}^{(K)})$.

So we just minimize the energy, because that would maximize the probability density. Clearly, there are many ways to implement (2a), e.g. gradient descent, conjugate gradient, Newton methods etc. (see Numerics II).

In complex systems, the energy landscape $\Phi(\mathbf{r})$ will have many local minima, so the result will depend very much on the initial configuration $\mathbf{r}^{(0)}$, and we will usually have to repeat this algorithm many times with different initial configurations in order to find a good energy minimum. Even then, this algorithm becomes in practice untractable for high-dimensional systems because they tend to have very rugged energy landscapes and an astronomical number of irrelevant local energy minima. Another way to deal with this problem may be to reduce the dimension, i.e. to replace the full (atomistic) coordinates $\mathbf{r}$ with some effective or coarse-grained coordinates in which the energy landscape is "smoother". Finding appropriate coarse-grained models for many-body problems in particular domains of science is a whole area of computational sciences by itself.

1.9 Computation

Before moving on, let us recap the essential computational steps necessary to simulate or optimize N-body systems using the methods above. We can summarize all algorithms that employ simple particle-pair potentials as follows:

1. Start with initial state $\mathbf{r}^{(0)}$ or $(\mathbf{r}^{(0)}, \mathbf{p}^{(0)})$.
2. For $k = 1, ..., K$:
   (a) Dynamics (stochastic or deterministic): Compute forces $\mathbf{f}$ with

$$\mathbf{f}_i(\mathbf{r}^{(k-1)}) = \nabla_i \Phi(\mathbf{r}^{(k-1)}) = \sum_j \mathbf{f}_{ji}(\mathbf{r}^{(k-1)})$$

   where $\mathbf{f}_{ij}(\mathbf{r})$ is the force from particle j onto particle i. Then update

$$\mathbf{r}^{(k)} = F\left\{\mathbf{r}^{(k-1)}, \mathbf{f}(\mathbf{r}^{(k-1)})\right\}$$

   (b) Monte-Carlo or Optimization: Compute energy $\Phi(\mathbf{r}^{(k)})$ with

$$\Phi(\mathbf{r}^{(k)}) = \sum_{\text{pairs } i,\, j} \mathbf{f}_{ji}(\mathbf{r}^{(k)})$$

   where $\mathbf{f}_{ij}(\mathbf{r})$ is the force from particle j onto particle i. Then update

$$\mathbf{r}^{(k)} = F\left\{\mathbf{r}^{(k-1)}, \Phi(\mathbf{r}^{(k)}), \Phi(\mathbf{r}^{(k-1)})\right\}$$

The main insight is that no matter which algorithm we will be using, and no matter if we need energies or forces, we will have to compute on the order of N^2 interaction terms between N particles. Since the number of time or iteration steps K can be huge, and the number of particles can be very large (e.g. $K = 10^{12}$ and $N = 10^6$ for molecular dynamics), the computational effort is going to be immense. In order to keep the computational cost under control, it is elementary that we use all possible tricks available to reduce this computational cost. In most of this lecture we will focus on methods to reducing the computational cost per iteration from $\mathcal{O}(N^2)$ to lower complexities, such as $\mathcal{O}(N^{3/2})$, $\mathcal{O}(N \ln N)$ or even $\mathcal{O}(N)$.

1.10 The four fundamental forces and interactions

In physics we generally distinguish four fundamental forces or interactions. Let's briefly review these before investigating some forces in more detail:

(taken from http://www.phy.duke.edu/~kolena/modern/forces.html)

Weak (nuclear) interaction:

The interaction range r can be roughtly estimated by assuming that during their lifetime τ particles move with 71% of the speed of light c in the laboratory frame (Lorentz factor $\gamma = 1.41$): $r \approx \gamma v \tau \approx c\tau$. With a lifetime of 3×10^{-25} s, this results in an interaction range of $0,09$ femtometer. A proton has a diameter of approximately one femtometer.

Strong (nuclear) interaction:

(taken from https://en.wikipedia.org/wiki/Strong_interaction)

In particle physics, the strong interaction is the mechanism responsible for the strong nuclear force (also called the strong force, nuclear strong force), and is one of the four known fundamental interactions, the others are electromagnetism, the weak interaction and gravitation. At the range of 10^{-15} m (femtometer), the strong force, is approximately 137 times stronger than electromagnetism, a million times stronger than the weak interaction and 1038 times stronger than gravitation.

The strong interaction is observable at two ranges: on a larger scale (about 1 to 3 femtometers (fm)), it is the force that binds protons and neutrons (nucleons) together to form the nucleus of an atom. On

the smaller scale (less than about 0.8 fm, the radius of a nucleon), it is the force (carried by gluons) that holds quarks together to form protons, neutrons, and other hadron particles.

Electromagnetic interaction:

Long-ranged, especially the charge-charge interaction. Can become short-ranged in the practical sense of computation (see below) if we consider special cases, such as the van der Waals force or higher multipole interactions. Can be both attractive and repulsive

Gravity:

Long-ranged, but of weak absolute magnitude. If bodies are charged, will be dominated by electromagnetic interactions. Is always attractive.

2 Electrodynamics

2.1 Notation

Taken from Eric Poisson's lecture notes.

Maxwell's classical theory of the electromagnetic field involves the vector fields:

$\mathbf{E}(\mathbf{r}, t)$	electric field at postition $\mathbf{x}$ and time t
$\mathbf{B}(\mathbf{r}, t)$	magnetic field at postition $\mathbf{x}$ and time t.

Thus we have two vectors at each position of space and time. In contrast to classical mechanics, where we usually operator with a finite number of objects, our field variables have in principle infinitely many degrees of freedom.

Electric and magnetic fields are produced by charges and currents. An *element of charge* is a small portion of matter that contains a net charge. In continuous electrodynamics theory this element must be sufficiently large to contain a macroscopic number of elementary charges, but sufficiently small that the density of charge within the volume is approximately uniform. This element is described by the following quantities:

$\rho(\mathbf{r}, t)$	charge density at postition $\mathbf{x}$ and time t
$\mathbf{v}(\mathbf{r}, t)$	velocity of element of charge at $\mathbf{x}$ and t.
$\mathbf{j}(\mathbf{r}, t)$	current density at $\mathbf{x}$ and t.

and note the important relation

$$\mathbf{j} = \rho\mathbf{v}$$

We use the following mathematical notation: The nabla operator $\boldsymbol{\nabla}$ is (in a canonical coordinate system of x, y, z):

$$\boldsymbol{\nabla} \simeq \begin{pmatrix} \nabla_x \\ \nabla_y \\ \nabla_z \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

and when applied to a scalar field $f(\mathbf{r})$ gives the gradient

$$\nabla f = \text{grad } f \simeq \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}$$

and similarly applied to a vector field $\mathbf{A}(\mathbf{r})$ we compute the scalar divergence

$$\nabla\mathbf{A} = \text{div } \mathbf{A} \simeq \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \mathbf{A} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} = \left(\frac{\partial}{\partial x} A_x + \frac{\partial}{\partial y} A_y + \frac{\partial}{\partial z} A_z \right)$$

For other coordinate systems these expressions change and we will introduce these as needed. If you are interested more in analysis on fields you might consider a lecture in Vector Analysis or for the more general case Tensor Analysis/Differential Geometry.

The Laplace Δ operator is given by:

$$\Delta \equiv \nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The gradient of a point-distance from $\mathbf{r}$ to origin $\mathbf{r}_0$ is

$$\boldsymbol{\nabla}|\mathbf{r} - \mathbf{r}_0| = \frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|}. \tag{3}$$

2.2 Maxwell's equation and the Lorentz force

The four Maxwell equations determine the electromagnetic field once the charge and current distributions are specified:

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \mathbf{E}(\mathbf{r},\, t) &= \frac{1}{\epsilon_0}\rho \\
\boldsymbol{\nabla} \cdot \mathbf{B}(\mathbf{r},\, t) &= 0 \\
\boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r},\, t) &= -\frac{\partial \mathbf{B}(\mathbf{r},\, t)}{\partial t} \\
\boldsymbol{\nabla} \times \mathbf{B}(\mathbf{r},\, t) &= \mu_0 \mathbf{j}(\mathbf{r},\, t) + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}(\mathbf{r},\, t)}{\partial t}
\end{aligned}$$

Where ϵ_0 and μ_0 are constants. The Maxwell equations state that:

1. the electric field is produced by charges (Eq. 1) and time-varying magnetic fields (Eq. 3)
2. the magnetic field is produced by currents and time-varying electric fields (Eq. 4)

Maxwell's equations can also be written in integral form, by involking the Gauss and Stokes theorems of vector calculus.

Since the electric and magnetic field act upon charged particles these will alter their positions and velocities: The Lorentz force determines the motion of the charges once the electromagnetic field is specified. Let:

$$\mathbf{f}(\mathbf{r},\, t) \qquad \text{force density at postition } \mathbf{r} \text{ and time } t$$

where the force density is defined as the net force acting on an element of charge at $\mathbf{r}$ divided by the volume of this charge element. The statement of the Lorentz force is:

$$\begin{aligned}
\mathbf{f}(\mathbf{r},\, t) &= \rho(\mathbf{r})\mathbf{E}(\mathbf{r},\, t) + \mathbf{j}(\mathbf{r},\, t) \times \mathbf{B}(\mathbf{r},\, t) \\
&= \rho(\mathbf{r}) \left(\mathbf{E}(\mathbf{r},\, t) + \mathbf{v}(\mathbf{r},\, t) \times \mathbf{B}(\mathbf{r},\, t)\right).
\end{aligned}$$

The net force $\mathbf{F}_V$ acting on a volume V of the charge distribution is the integral over its force densities:

$$\mathbf{F}_V = \int_V \mathbf{f}(\mathbf{r},\, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$$

For a single point charge at position $\mathbf{r}(t)$ we have

$$\begin{aligned}
\rho(\mathbf{x},\, t) &= q\delta(\mathbf{x} - \mathbf{r}(t)) \\
\mathbf{j}(\mathbf{x},\, t) &= q\mathbf{v}\delta(\mathbf{x} - \mathbf{r}(t))
\end{aligned}$$

where we used the Dirac delta distribution (note that this is not really a function, but we treat it as such). With this the total force $\mathbf{F}_{pc}$on the point charge becomes:

$$\mathbf{F}_{pc}(t) = q\left(\mathbf{E}(\mathbf{r}(t)) + \mathbf{v} \times \mathbf{B}(\mathbf{r}(t))\right)$$

Here the fields are evaluated at the current charge's position.

Taken together, the Maxwell equations and the Lorentz force determine the behavior of the charge and current distributions, and the time evolution of the electric and magnetic fields. These five equations summarize the complete theory of classical electrodynamics. Every conceivable phenomenon involving electromagnetism can be predicted from them.

The remaining content of electrodynamics lectures is the study of consequences and special cases of the above equations. In particular, we have the following fundamental consequences:

1. Charge conservation - we can move charges around, but they are not destroyed or created.
2. Energy conservation - total energy is conserved
3. Momentum conservation - total momentum is conserved when the field evolves

2.3 Potentials

Maxwell's equations hint at the fact that the electric and magnetic fields are not independent and we can use this to express the electric and magnetic field by *smaller* objects, a magnetic vector potential $\phi(\mathbf{r}, t)$ and a scalar electric potential $\Phi(\mathbf{x}, t)$

$$\mathbf{B}(\mathbf{r}, t) \equiv \nabla \times \mathbf{A}(\mathbf{r}, t)$$
$$\mathbf{E}(\mathbf{r}, t) \equiv -\nabla\phi - \frac{\partial}{\partial t}\mathbf{A}(\mathbf{r}, t)$$

What matters for our description of the physics is the resulting $\mathbf{B}$ and $\mathbf{E}$ field and hence any transformation of $\mathbf{A}$ and ϕ that leaves the physics unchanged is an equally good representation of the state of the system. Since we are free of this choice of a transformation we can pick a representation of $\mathbf{A}$ and ϕ that suits us best. This is similar to our freedom of choice of the coordinate system, e.g. use spherical coordinates of radial symmetric system because it simplifies equations and the physics is the same regardless of our choice of coordinates. In our case the particular choice is called a gauge.

Consider the transformation where we add a vector field that is generated as the gradient of a scalar function ψ

$$\mathbf{A}(\mathbf{r}, t) \rightarrow \mathbf{A}(\mathbf{r}, t) + \nabla\psi(\mathbf{r}, t)$$

then $\mathbf{B}$ does not change since $\nabla \times (\nabla\psi) = 0$ for any scalar field ψ provided it can be differentiated twice. Similarly we need to transform the scalar potential ϕ by

$$\varphi(\mathbf{r}, t) \rightarrow \varphi(\mathbf{r}, t) - \frac{\partial}{\partial t}\psi(\mathbf{r}, t)$$

It is possible to find a potential ψ so that the so-called *Lorenz gauge* condition

$$\nabla\mathbf{A} + \frac{1}{c}\frac{\partial}{\partial t}\phi = 0$$

holds using which we can (proof now shown) rewrite Maxwell's equations into the following wave equations:

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\Phi(\mathbf{x}, t) = -\frac{1}{\epsilon_0}\rho(\mathbf{x}, t)$$
$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\mathbf{A}(\mathbf{x}, t) = -\mu_0\mathbf{j}(\mathbf{x}, t)$$

where $c = 1/\sqrt{\epsilon_0\mu_0}$ is the speed of light in vacuum.

2.4 Time-stationary solution

In time-independent solutions, the fields and potentials no longer depend on t, so all derivatives w.r.t. time disappear $\partial_t \equiv 0$ and the equations reduce to:

$$\nabla^2\Phi(\mathbf{x}) = -\frac{1}{\epsilon_0}\rho(\mathbf{x})$$
$$\nabla^2\mathbf{A}(\mathbf{x}) = -\mu_0\mathbf{j}(\mathbf{x})$$

which are Poisson's equations and we can employ the full theory on Poisson's equation to solve stationary problems.

Note, that now $\mathbf{E} = -\boldsymbol{\nabla}\Phi$ and $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ are independent and the *Lorenz gauge* condition reduces to the *Coulomb gauge* condition

$$\boldsymbol{\nabla} \cdot \mathbf{A} = 0$$

3 Electrostatics

3.1 Basic equations

Electrostatics denotes the time-stationary (and hence time-independent) solution of the electric field only. Poisson's equation for the electric potential (as briefly derived in chapter 2) is

$$\Delta\Phi(\mathbf{r}) = \nabla^2\Phi(\mathbf{r}) = -\frac{1}{\epsilon_0}\rho(\mathbf{r})$$

together with the potential-field relation

$$\mathbf{E} = -\boldsymbol{\nabla}\Phi$$

to compute the stationary electric field once the potential is known. The Poisson equation relates the charge density and the potential field generated by it. For special geometries the Poisson equation can be solved directly while the general solution to Poisson's equation can be written using the *Green's function* approach

$$\Phi(\mathbf{r}) = \Phi_0(\mathbf{r}) + \frac{1}{4\pi\epsilon_0}\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}' \tag{4}$$

The idea of the Green's function in a nutshell: We want to compute the potential in Laplace's equation with a given charge distribution. You could do this by integration directly which can be seen if you consider only one dimension. But we take a different approach and exploit that for a linear differential operator such as the Laplace operator the superposition principle holds. So, if we would know the solution for a point charge at $\mathbf{r}'$ called the Greens function G

$$\Delta G_{\mathbf{r}'}(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}')$$

distribution we could combine these solutions to solve Laplace's equation for an arbitrary charge distribution. We integrate over the charge density times the point solution

$$\int d\mathbf{r}'\,\rho(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}') = \int d\mathbf{r}'\,\rho(\mathbf{r}')\Delta G_{\mathbf{r}'}(\mathbf{r})$$

$$\rho(\mathbf{r}) = \Delta\int d\mathbf{r}'\,\rho(\mathbf{r}')G_{\mathbf{r}'}(\mathbf{r})$$

and so

$$\phi(\mathbf{r}) = \int d\mathbf{r}'\,\rho(\mathbf{r}')G_{\mathbf{r}'}(\mathbf{r})$$

is the solution. Note, that the Green's function only depends on the differential operator. For the Laplace operator Δ the solution is (without derivation)

$$\Delta \leftrightarrow G_{r'}(r) = \frac{1}{4\pi|\mathbf{r}-\mathbf{r}'|}$$

which leads to the final eq. 4.

The $\Phi_0(\mathbf{r})$ is the constant of integration and is any solution that fulfills the homogeneous Laplace equation

$$\Delta\phi(\mathbf{r}) = 0$$

used to enforce boundary conditions that we might wish to impose on the scalar potential.

3.2 Gaussian charge density

We will try to derive the single point charge by integration now. Consider first the example of a Gaussian charge density ρ_σ^G with charge q centered at the origin and variance σ^2

$$\rho_\sigma^G(r) = \frac{q}{\sigma^3(2\pi)^{3/2}}\exp\left(-\frac{r^2}{2\sigma^2}\right)$$

where $r = |\mathbf{r}|$ is the distance from the origin. Since this is a radial symmetric we use the representation of $\Delta = \nabla^2$ in spherical coordinates and know that only $\partial_r \neq 0$:

$$\Delta_r \phi_\sigma(r) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\phi_\sigma(r)\right) = -\frac{1}{\epsilon_0}\rho_\sigma^G(r)$$

integrating two times with respect to r and rearranging yields the solution

$$\phi_\sigma(r) = \frac{q}{4\pi\epsilon_0 r}\mathrm{erf}\left(\frac{r}{\sqrt{2}\sigma}\right) \tag{5}$$

where erf is the so-called *error function,* the antiderivative of the Gauss' function. Taking the limit $\sigma \to 0$ provides the potential of a point charge at the origin

$$\phi_\delta(r) = \frac{q}{4\pi\epsilon_0 r}$$

which the characteristic $1/r$ dependence on the distance. Comparison yields

$$\Delta\frac{q}{4\pi\epsilon_0 r} = \Delta\phi_\delta(r) = \frac{1}{\epsilon_0}\rho_\sigma^G(r) = \frac{1}{\epsilon_0}q\delta(r) = \frac{1}{\epsilon_0}q\Delta G_0(r)$$

the Greensfunktion for the charge at the origin and similarly the solution for an arbitrary position $\mathbf{r}'$.

3.3 Point charge

In general we have several point charges and let i reference point charge q_i at position $\mathbf{r}_i$:

$$\rho_i(\mathbf{r}) = q_i\delta(\mathbf{r} - \mathbf{r}_i).$$

The electric potential generated by one of the point charge is

$$\phi_i(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}_i|} \tag{6}$$

Using (3), we can also write the electric field:

$$\mathbf{E}_i(\mathbf{r}) = -\boldsymbol{\nabla}\Phi(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0}\frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3}.$$

Thus the electric field has a magnitude proportional to inverse distance squared and points towards positive charges and away from negative charges. This well-known result is known as *Coulomb's law.*
We generalize (6) by calling $\rho_i(\mathbf{r})$ be an arbitrary charge density of ion i, not necessarily the distribution of a point charge or a Gaussian distribution. Remember that we can write the potential of an arbitrary charge density as

$$\phi_i(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{r}'\,\frac{\rho_i(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \tag{7}$$

Integrating the Lorentz force along the path of a point charge $q(\mathbf{r})$ yields the change in electrostatic energy. Since the potential is the anti-derivative of the electric field $\mathbf{E} = -\nabla\phi$ it turns out that the electrostatic interaction energy of a point charge with the potential field $\phi(\mathbf{r})$ is given by:

$$E = q(\mathbf{r})\phi(\mathbf{r})$$

where we assumed that for $\mathbf{r} \to \infty$ the energy vanishes. In the special case of two point charges q_i and q_j at positions $\mathbf{r}_i$ and $\mathbf{r}_j$ the interaction energy is :

$$E_{ij} \equiv \frac{q_i q_j}{4\pi\epsilon_0|\mathbf{r}_j - \mathbf{r}_i|}.$$

Finally, the total interaction energy E of N charges $q_1, ..., q_N$ is given by:

$$E = \sum_{i>j}\frac{q_i q_j}{4\pi\epsilon_0|\mathbf{r}_j - \mathbf{r}_i|} = \frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{4\pi\epsilon_0|\mathbf{r}_j - \mathbf{r}_i|}$$

Note, that the interaction energy is only counted once per pair. The interaction energy E_{ij}can be considered as the energy of bringing particle i to distance r of particle j.

3.4 Dipole

Another elementary situation is that of two equal charges, one positive, one negative, separated by a distance $d = 2\epsilon$. We align the charges along the direction of the unit vector $\hat{\mathbf{p}}$ and place the origin of the coordinate system at the center point.

The charge density is then given by

$$\rho(\mathbf{r}) = q\delta(\mathbf{r} - \epsilon\hat{\mathbf{p}}) - q\delta(\mathbf{r} + \epsilon\hat{\mathbf{p}})$$

and the potential is:

$$\Phi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{|\mathbf{r} - \epsilon\hat{\mathbf{p}}|} - \frac{1}{|\mathbf{r} + \epsilon\hat{\mathbf{p}}|}\right).$$

We can simplify this expression if we take the observation point $\mathbf{x}$ to be far away from the dipole:

$$r = |\mathbf{r}| \gg \epsilon$$

Keeping terms of order ϵ only, the potential becomes

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{\langle 2q\epsilon\hat{\mathbf{p}},\, \mathbf{r}\rangle}{r^3}.$$

The vector $\mathbf{p} = 2q\epsilon\hat{\mathbf{p}}$ is the *dipole moment* or the charge distribution. In general, for a collection of several charges, this is defined as

$$\mathbf{p} = \sum_i q_i \mathbf{r}_i.$$

In the present situation the definition implies $\mathbf{p} = (+q)(\epsilon\hat{\mathbf{p}}) + (-q)(-\epsilon\hat{\mathbf{p}}) = 2q\epsilon\hat{\mathbf{p}}$. The potential of the dipole is

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{\langle \hat{\mathbf{p}},\, \mathbf{r}\rangle}{r^3}$$

when $r \gg \epsilon$. Notice that the potential falls of as $1/r^2$, faster than for a single point charge. This is due to the fact that the charge distribution has a vanishing total charge.

3.5 Numbers

TIP3P water model:

r(OH)	0.9572 A	0.09572 nm
$\angle$(HOH)	104.52	104.52
A	$582\,10^{-3}\frac{kcal\,A^{12}}{mol}$	$582\,10^{-3} \times 4.184\frac{kJ}{kcal}\frac{kcal\,10^{-12}nm^{12}}{mol} = 2.435088\,10^{-12}\frac{kJ\,nm^{12}}{mol}$
B	$595\,\frac{kcal\,A^6}{mol}$	$595 \times 4.184\frac{kJ}{kcal}\frac{kcal\,10^{-6}nm^6}{mol} = 2.48949 \times 10^{-3}\frac{kJ\,nm^{12}}{mol}$
q(O)	-0.834	-0.834
q(H)	0.417	0.417

Elementary charge: $q_0 = 1.60217656535 \times 10^{-19}C$

Dielectric constant of vacuum: $\epsilon_0 = 8.854187817 \times 10^{-12}\frac{C^2}{Nm^2}$

Distance in Angstrom: $r_{ij}[A] = 10^{-10}m$

Avogadro constant $N = 6.0221413 \times 10^{23}$

$$\begin{aligned} E_{ij} &= \frac{1}{r/nm}\frac{(1.60217656535 \times 10^{-19}C)^2 \times 6.0221413 \times 10^{23}\frac{1}{mol}}{4\pi \times 8.854187817 \times 10^{-12}\frac{C^2}{Jm}10^{-9}m}10^{-3}\frac{kJ}{J} \\ &= 86.716695983218\frac{kJ}{mol} \end{aligned}$$

Density of liquid water:

$$33.3679 \times 10^{27} m^{-3} = 33.3679\, nm^{-3}$$

Density of ice

$$30.5944 \times 10^{27} m^{-3} = 30.5944\, nm^{-3}$$

4 Short-ranged forces

4.1 Van der Waals forces

[**From Van der Waals Wiki**]: Van der Waals forces include attraction and repulsions between atoms, molecules, and surfaces, as well as other intermolecular forces. They differ from covalent and ionic bonding in that they are caused by correlations in the fluctuating polarizations of nearby particles (a consequence of quantum dynamics[3]).

Intermolecular forces have four major contributions:

A repulsive component resulting from the Pauli exclusion principle that prevents the collapse of molecules. Attractive or repulsive electrostatic interactions between permanent charges (in the case of molecular ions), dipoles (in the case of molecules without inversion center), quadrupoles (all molecules with symmetry lower than cubic), and in general between permanent multipoles. The electrostatic interaction is sometimes called the Keesom interaction or Keesom force after Willem Hendrik Keesom. Induction (also known as polarization), which is the attractive interaction between a permanent multipole on one molecule with an induced multipole on another. This interaction is sometimes called Debye force after Peter J.W. Debye. Dispersion (usually named after Fritz London), which is the attractive interaction between any pair of molecules, including non-polar atoms, arising from the interactions of instantaneous multipoles.

[**From Lennard-Jones Wiki**]: The Lennard-Jones potential[1] (also termed the L-J potential, 6-12 potential, or 12-6 potential) is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones.[1] The most common expressions of the L-J potential are:

$$V_{\mathrm{LJ}} = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] = \varepsilon \left[\left(\frac{r_{\mathrm{m}}}{r}\right)^{12} - 2\left(\frac{r_{\mathrm{m}}}{r}\right)^{6} \right],$$

where ϵ is the depth of the potential well, σ is the finite distance at which the inter-particle potential is zero, r is the distance between the particles, and rm is the distance at which the potential reaches its minimum. At rm, the potential function has the value $-\epsilon$. The distances are related as $r_m = 21/6\sigma \approx 1.122\sigma$. These parameters can be fitted to reproduce experimental data or accurate quantum chemistry calculations. Due to its computational simplicity, the Lennard-Jones potential is used extensively in computer simulations even though more accurate potentials exist.

5 Gravity

Taken from https://en.wikipedia.org/wiki/Gravitational_potential

Before going on with forces in microscopic systems, it is worth considering another force that is very different from a physical point of view, but very similar from a mathematical point of view. The latter similarity leads to common ground in the mathematical, numerical and algorithmic treatment.

We will study the phenomenon of gravity in the limit of classical mechanics. Note that the correct description of masses moving at high speeds and the description of the interaction of masses and electromagnetic waves requires a relativistic treatment of the matter which is beyond our current scope.

[1] Lennard-Jones, J. E. (1924), "On the Determination of Molecular Fields", Proc. R. Soc. Lond. A, 106 (738): 463–477

The **gravitational potential** V is the graviational potential energy U per unit mass:

$$U = mV$$

where m is the mass of the object. The potential energy is equal (in magnitude, but negative) to the work done by the gravitational field moving a body to its given position in space from infinity. If the body has a mass of 1 unit, then the potential energy to be assigned to that body is equal to the gravitational potential. So the potential can be interpreted as the negative of the work done by the gravitational field moving a unit mass in from infinity. The graviational potential is analogous to the electric potential with mass playing the role of charge. **TODO: relate to equations**.

In some situations, the equations can be simplified by assuming a field that is nearly independent of position. For instance, in daily life, in the region close to the surface of the Earth, the gravitational acceleration can be considered constant. In that case, the difference in potential energy from one height to another is to a good approximation linearly related to the difference in height:

$$\Delta U = mg\Delta h$$

where g is the gravitational constant

Gravitational force

From: https://en.wikipedia.org/wiki/Gravity

In 1687, English mathematician Sir Isaac Newton published Principia, which hypothesizes the inverse-square law of universal gravitation. In his own words, "I deduced that the forces which keep the planets in their orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the Earth; and found them answer pretty nearly."[3] The equation is the following:

$$F = G\frac{m_1 m_2}{r^2}$$

Where F is the force, m_1 and m_2 are the masses of the objects interacting, r is the distance between the centers of the masses and G is the gravitational constant.

Speed of gravity: In classical theories of gravitation, the changes in a gravitational field propagate. A change in the distribution of energy and momentum of matter results in subsequent alteration, at a distance, of the gravitational field which it produces. In a more physically correct sense, the "speed of gravity" refers to the speed of a gravitational wave, which is the same speed as the speed of light (c).

Mathematical form

The potential V of a unit mass m at a distance r from a point mass of mass M can be defined as the work W done by the gravitational force F bringing the unit mass in from infinity to that point:[1][2][3]

$$V(r) = \frac{W}{m} = \frac{1}{m}\int_{\infty}^{r} F\,\mathrm{d}r = \frac{1}{m}\int_{\infty}^{r} \frac{GmM}{r^2}\,\mathrm{d}r = -\frac{GM}{r}.$$

where G is the universal gravitational constant with a value of approximately $6.674 \times 10^{-11}\,\mathrm{Nm^2/kg^2}$. The potential has units of energy per unit mass, e.g., J/kg. By convention, it is always negative where it is defined, and converges to zero with $x \to \infty$.

The gravitational field, and thus the acceleration of a small body in the space around the massive object, is the negative gradient of the gravitational potential. Thus the negative of a negative gradient yields positive acceleration toward a massive object. Because the potential has no angular components, its gradient is

$$\mathbf{a} = -\frac{GM}{r^3}\mathbf{r} = -\frac{GM}{r^2}\hat{\mathbf{r}},$$

where $\mathbf{r}$ is a vector of length r pointing from the point mass toward the small body and $\hat{\mathbf{r}}$ is a unit vector pointing from the point mass toward the small body. The magnitude of the acceleration therefore follows an inverse square law:

$$|\mathbf{a}| = \frac{GM}{x^2}.$$

The potential associated with a mass distribution is the superposition of the potentials of point masses. If the mass distribution is a finite collection of point masses, and if the point masses are located at the points $\mathbf{x}_1$, ..., $\mathbf{x}_n$ and have masses m_1, ..., m_n, then the potential of the distribution at the point $\mathbf{x}$ is

$$V(\mathbf{r}) = -\sum_{i=1}^{n} \frac{Gm_i}{|\mathbf{r} - \mathbf{r_i}|}.$$

If the mass distribution is given as a mass measure $\mathrm{d}m$ on three-dimensional Euclidean space $\mathbb{R}^3$, then the potential is the convolution of $-G/|\mathbf{r}|$ with $\mathrm{d}m$. In good cases this equals the integral

$$V(\mathbf{r}) = -\int_{\mathbf{R}^3} \frac{G}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}m(\mathbf{r}'),$$

If there is a function $\rho(\mathbf{r})$ representing the density of the distribution at $\mathbf{r}$, so that $\mathrm{d}m(\mathbf{r}) = \rho(\mathbf{r})\, \mathrm{d}v(\mathbf{r})$, where $\mathrm{d}v(\mathbf{r})$ is the Euclidean volume element, then the gravitational potential is the volume integral

$$V(\mathbf{r}) = -\int_{\mathbf{R}^3} \frac{G}{|\mathbf{r} - \mathbf{r}'|}\, \rho(\mathbf{r}')\, \mathrm{d}v(\mathbf{r}').$$

If V is a potential function coming from a continuous mass distribution $\rho(\mathbf{r})$, then ρ can be recovered using the Laplace operator, Δ:

$$\rho(\mathbf{r}) = \frac{1}{4\pi G} \Delta V(\mathbf{r}).$$

This holds pointwise whenever ρ is continuous and is zero outside of a bounded set. In general, the mass measure dm can be recovered in the same way if the Laplace operator is taken in the sense of distributions. As a consequence, the gravitational potential satisfies Poisson's equation. See also Green's function for the three-variable Laplace equation and Newtonian potential.

Numerical Values

The absolute value of gravitational potential at a number of locations with regards to the gravitation from [clarification needed] the Earth, the Sun, and the Milky Way is given in the following table; i.e. an object at Earth's surface would need 60 MJ/kg to "leave" Earth's gravity field, another 900 MJ/kg to also leave the Sun's gravity field and more than 130 GJ/kg to leave the gravity field of the Milky Way. The potential is half the square of the escape velocity.

Location	W.r.t. Earth	W.r.t. Sun	W.r.t. Milky Way
Earth's surface	60 MJ/kg	900 MJ/kg	≥130 GJ/kg
LEO	57 MJ/kg	900 MJ/kg	≥130 GJ/kg
Voyager 1 (17×10^9 km from Earth)	23 J/kg	8 MJ/kg	≥130 GJ/kg
0.1 light-year from Earth	0.4 J/kg	140 kJ/kg	≥130 GJ/kg

6 Particle-particle methods

Suppose that we have a system with N particles and the interaction energy between particles i and j is given by $e(r_{ij})$, where $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ is the interparticle distance. The direct approach to compute interactions is:

$E \leftarrow 0$

`For` $i = 1, ..., N-1$:

  $E_i \leftarrow 0$

  `For` $j = i+1, ..., N$:

    $E_i \leftarrow E_i + e(r_{ij})$

  $E \leftarrow E + E_i$

This algorithm has obviously a computational complexity of $\mathcal{O}(N^2)$, with $\frac{N(N-1)}{2}$ interactions to be computed. For large N this approach becomes too expensive.

6.1 Cutoff

A simple idea to save costs is the following: Since all interaction energies (and consequently also the forces) decay as a function of r we have

$$\lim_{r_{ij}\to\infty} e(r_{ij}) = 0$$

Thus we pursue the simple idea of ignoring small interaction energies at long distances. The simplest version of that idea is to use a cutoff:

$$e^{\text{cut}}(r_{ij}) = \begin{cases} e(r_{ij}) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c. \end{cases}$$

However, this function is discontinuous and not differentiable at $r_{ij} = r_c$. For this reason it is not suitable for computing particle dynamics. A simple improvement is to use a shift:

$$e^{\text{shift}}(r_{ij}) = \begin{cases} e(r_{ij}) - e(r_c) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c. \end{cases}$$

Which changes the interaction energy by a constant, but this has no effect on the force. A shifted-cutoff potential still has a kink at $r_{ij} = r_c$, and thus cannot be differentiated twice (to compute the accelerations). Thus, in practice one prefers to use a switch function:

$$e^{\text{switch}}(r_{ij}) = \begin{cases} e(r_{ij}) & r_{ij} < r_{\text{o}} \\ s\left(\frac{r_{ij}-r_{\text{o}}}{r_{\text{c}}-r_{\text{o}}}\right) e(r_{ij}) & r_{\text{o}} \leq r_{ij} < r_{\text{c}} \\ 0 & r_{\text{c}} \leq r_{ij}. \end{cases}$$

where $s(x)$ is a function that smoothly decays from $s(0) = 1$ to $s(1) = 0$ and is twice differentiable in $x \in [0, 1]$. Different choices can be made here.

6.2 Cost

Suppose for simplicity that our particle system lives in a volume V and we have an approximately uniform particle density $\rho = N/V$. We also suppose that there is a gived maximum density, and regard ρ to be a constant, i.e. if we want to change N, we must also increase V accordingly. We work in 3 dimensions, but results can be adapted for other dimensions.

When using a cutoff scheme, each particle has a neighborhood volume of $v = \frac{4}{3}\pi r_c^3$ and thus on average $n = \frac{4}{3}\pi r_c^3 \rho$ neighbors. Note that by convention ρ is a constant, i.e. independent of N. Each particle interaction only needs to be computed once, thus once we have the neighborhoods, we have to compute $N\frac{n-1}{2}$ interactions, and the complexity has dropped to $\mathcal{O}(N)$.

Note that in order to be better than $\mathcal{O}(N^2)$ in the overall computation we also have to use a nontrivial way to compute the neighborhoods, because direct computation of all pair distances is $\mathcal{O}(N^2)$ itself (to be discussed later).

Note further that in order to get a good accuracy, we might have to choose r_c large, which produces a large prefactor in the computational cost.

6.3 Error

Energy error with uniform interaction strengths: We consider a system of N interacting bodies at uniform density in space. For simplicity we start by a situation where the interaction strength is uniform, e.g. gravitational interaction between equal masses. This can be expressed by the interaction law

$$e(r_{ij}) = \frac{a}{r_{ij}^p}$$

We first use the uniform particle distribution:

$$n(r)\mathrm{d}r = 4\pi\rho r^2 \mathrm{d}r$$

with particle density ρ (Number of particles per unit volume). We confirm that this gives:

$$N(R) = 4\pi\rho \int_{r=0}^{R} r^2 dr = 4\pi\rho[\frac{1}{3}r^3]_0^R = \rho\frac{4}{3}\pi R^3 = \rho V_R.$$

We write the total interaction energy between a particle in the origin and all particles *outside* a radius R in three dimensions as

$$\begin{aligned} e_R &= \int_{r=R}^{R_{max}} n(r)\frac{a}{r^p}\mathrm{d}r \\ &= 4\pi\rho a \int_{r=R}^{R_{max}} \frac{1}{r^{p-2}}\mathrm{d}r \\ &= -\frac{4\pi\rho a}{p-3}\left[\frac{1}{r^{p-3}}\right]_R^{R_{max}} \end{aligned}$$

and thus

$$\begin{aligned} e_R &= \frac{4\pi\rho a}{p-3}\left[\frac{1}{R^{p-3}} - \frac{1}{R_{max}^{p-3}}\right] \\ E &= \frac{4\pi\rho a}{p-3}\left[\frac{1}{R_{min}^{p-3}} - \frac{1}{R_{max}^{p-3}}\right] \end{aligned}$$

Case 1: long-ranged interactions: For $p < 3$, we have

$$e_R = \frac{4\pi\rho a}{p-3}\left[R^{3-p} - R_{max}^{3-p}\right]$$

parametrizing $R = \alpha R_{max}$

$$e_R = \frac{4\pi\rho a}{p-3}R_{max}^{3-p}\left[\alpha^{3-p} - 1\right]$$

shows that we have a system-size dependent error up to all $\alpha < 1$. In particular, for $R_{max} \to \infty$, we have

$$\lim_{R_{max}\to\infty} e_R = \infty \quad \forall R < \infty.$$

The relative error has the behavior

$$\frac{e_R}{E} = \frac{R^{3-p} - R_{max}^{3-p}}{R_{min}^{3-p} - R_{max}^{3-p}} = \left[\alpha^{3-p} - 1\right]\frac{R_{max}^{3-p}}{R_{min}^{3-p} - R_{max}^{3-p}}$$

For $R_{max} \gg R_{min}$

$$\frac{e_R}{E} \approx 1 - \alpha^{3-p}$$

and for $R_{max} \to \infty$

$$\lim_{R_{max}\to\infty} \frac{e_R}{E} = \lim_{R_{max}\to\infty} \frac{R^{3-p}}{R_{min}^{3-p} - R_{max}^{3-p}} - \lim_{R_{max}\to\infty} \frac{R_{max}^{3-p}}{R_{min}^{3-p} - R_{max}^{3-p}} = 1$$

Case 2: $p = 3$: We write

$$e_R = 4\pi\rho a \left[\frac{1}{(p-3)R^{p-3}} - \frac{1}{(p-3)R_{max}^{p-3}} \right]$$
$$E = 4\pi\rho a \left[\frac{1}{(p-3)R_{min}^{p-3}} - \frac{1}{(p-3)R_{max}^{p-3}} \right]$$

we use the limit

$$\lim_{q\to 0^+} \left[\frac{1}{qa^q} - \frac{1}{qb^q} \right] = \log \frac{b}{a}$$

and thus

$$e_R = 4\pi\rho a \log \frac{R_{max}}{R}$$
$$E = 4\pi\rho a \log \frac{R_{max}}{R_{min}}$$

Thus if we parametrize $R = \alpha R_{max}$, we get

$$e_R = -4\pi\rho a \log \alpha$$

We thus retain a finite but system-size independent error. The relative error

$$\frac{e_R}{E} = \frac{\log R_{max} - \log R}{\log R_{max} - \log R_{min}}$$

converges for $R \to R_{min}$, but slowly.

Case 3: short-ranged interactions with $p > 3$:

$$e_R = \frac{4\pi\rho a}{p-3} \left[\frac{1}{R^{p-3}} - \frac{1}{R_{max}^{p-3}} \right]$$

and we have a proper limit with

$$\lim_{R\to R_{max}} e_R = 0$$

In particular for $R_{max} \to \infty$

$$\lim_{R_{max}\to\infty} e_R = \frac{4\pi\rho a}{p-3} \frac{1}{R^{p-3}}$$

We can see that the error decays to 0.

The relative error is

$$\frac{e_R}{E} = \frac{\frac{1}{R^{p-3}} - \frac{1}{R_{max}^{p-3}}}{\frac{1}{R_{min}^{p-3}} - \frac{1}{R_{max}^{p-3}}}$$

and for $R_{max} \to \infty$

$$\lim_{R_{max}\to\infty} \frac{e_R}{E} = \frac{R_{min}^{p-3}}{R^{p-3}}$$

thus if we parametrize $R = \alpha R_{min}$, we can see that

$$\lim_{R_{max} \to \infty} \frac{e_R}{E} = \alpha^{3-p}$$

decays at least with α^{-1} or faster.

Thus we need $p \geq 3$ in order to be able to use a cutoff. For $p < 3$, the energy acting on a particle doesn't decay as a function of r_{ij}. As a result we can distinguish long-ranged forces, for which cutoffs are not suitable and short-ranged forces which can be effectively cut off. In $\mathbb{R}^3$ we have:

long-ranged	short-ranged
Coulomb interaction	Dipole-Dipole and higher multipole interactions
Charge-dipole interaction	Debye-Hückel (screened electrostatics)
Gravity	Van der Waals / Lennard Jones

Note that this classification depends on the space dimension. For example, in $\mathbb{R}^2$, the charge-dipole interaction would be short-ranged.

7 Neighbor search

[TODO]

8 Direct particle-field methods

8.1 Potential, interaction energy, and force

Example: Gravity and electrostatic (charge-charge) interaction in an N-body system. We can describe both effects by the potential and the respective force at a position $\mathbf{r} \in \mathbb{R}^3$ (**Note: beware of confusion - r was so far used for the full configuration vector of all particles - we should clear that up**):

$$\Phi(\mathbf{r}) = c \sum_{j=1}^{N} \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|},$$

$$\mathbf{f}(\mathbf{r}) = -\nabla\Phi(\mathbf{r}) = -c \sum_{j=1}^{N} \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|^2} \widehat{\mathbf{r} - \mathbf{r}_j}$$

With the choices:

Gravity: $c = G$ (gravitational constant), $q_j = m_j$ (mass of object)

Electrostatics: $c = 1/4\pi\epsilon_0$ (ϵ_0 dielectric coefficient of vacuum), q_j (charge)

8.2 Poisson equation

Let's go back to where these equations come from. We focus on electrostatics now, but as seen above the same rules work for gravity if we just exchange charges for masses and electrostatic for gravitational constants.

The electric field $\mathbf{E}$ due to a static charge density distribution $\rho(\mathbf{r})$ is determined by Gauss' Law (one of Maxwell's equations):

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0}$$

A static electric field can be derived from the electric potential $\phi(\mathbf{r})$: $\mathbf{E} = -\nabla\phi$, yielding the Poisson equation:

$$\nabla^2\phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0} \tag{8}$$

where the Laplace operator is given by

$$\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The Poisson equation relates the charge density and the potential field generated by it. For special geometries the Poisson equation can be solved directly. When considering a point charge at position $\mathbf{r}_j$, that means $\rho(\mathbf{r}) = \delta(|\mathbf{r} - \mathbf{r}_j|)$ the solution is (see earlier in the lecture notes)

$$\phi(\mathbf{r}) = \frac{q_j}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}_j|}$$

Which, for N charges leads to the equations above. For a continuum of chages we would have a superposition fields generated by point charges:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int d^3\mathbf{r}' \, \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

Now we take a different viewpoint than before: Instead of computing the interaction energies or forces between all pairs of charges in an N body system, we rather consider the interaction energy or force of a single charge at position $\mathbf{r}$ with the field generated by all other charges. If we have an N body system rather than a charge continuum it is at this point not obvious why this approach is useful, but we will see later that it is.

If we have the field $\phi(\mathbf{r})$, we can compute the electrostatic interaction energy and force of a point charge q_i at position $\mathbf{r}_i$ with a field $\phi(\mathbf{r})$ by:

$$e_i = q_i\phi(\mathbf{r}_i),$$
$$f_i = q_i\mathbf{f}(\mathbf{r}_i).$$

In order to work with this approach, we start by charge density $\rho(\mathbf{r})$ and then solve the Poisson equation in space in order to obtain $\phi(\mathbf{r})$. Thus, Eq. (8) is now the center of our attention. First we will discuss how to do this in principle, then we will discuss how to do it efficiently.

8.3 Uniqueness of the solution

(**This section is optional**)

Is the solution to Poisson's equation unique? I.e. if we consider a given charge density $\rho(\mathbf{r})$ and the two solutions.

$$\nabla^2\phi_1 = -\frac{\rho}{\epsilon_0}$$

and

$$\nabla^2\phi_2 = -\frac{\rho}{\epsilon_0}$$

is there a guarantee that $\phi_1 = \phi_2$ or at least a specific relation between them? This is a quite practically important question, because it means that if we find a solution ϕ_1 with some algorithm, for example an iterative algorithm that approaches the solution in steps and finally converges - do we know that the algorithm has found all relevant solutions, or would another algorithm, or the same algorithm starting from a different initial guess of the solution provide something else?

From the two equations above we get:

$$\nabla^2(\phi_1 - \phi_2) = 0$$

For any scalar s and any vector $\mathbf{v}$ we have the vector identity

$$\langle\nabla,\, s\mathbf{v}\rangle = s\langle\nabla,\, \mathbf{v}\rangle + \langle\mathbf{v},\, \nabla s\rangle$$

(**reference**). Selecting $\phi_1(\mathbf{r}) - \phi_2(\mathbf{r})$ as the scalar, and $\nabla(\phi_1(\mathbf{r}) - \phi_2(\mathbf{r}))$ for the vector for any given $\mathbf{r}$, we get:

$$\begin{aligned}\langle\nabla,\, (\phi_1 - \phi_2)\nabla(\phi_1 - \phi_2)\rangle =& (\phi_1 - \phi_2)\langle\nabla,\, \nabla(\phi_1 - \phi_2)\rangle \\ &+ \langle\nabla(\phi_1 - \phi_2),\, \nabla(\phi_1 - \phi_2)\rangle\end{aligned}$$

We will integrate this equation in volume, and if boundary surfaces are specified we will integrate within this volume:

$$\begin{aligned}\int_V \langle\nabla,\, (\phi_1 - \phi_2)\nabla(\phi_1 - \phi_2)\rangle dv =& \int_V (\phi_1 - \phi_2)\langle\nabla,\, \nabla(\phi_1 - \phi_2)\rangle dv \\ &+ \int_V \langle\nabla(\phi_1 - \phi_2),\, \nabla(\phi_1 - \phi_2)\rangle dv\end{aligned}$$

However,

$$\nabla,\, \nabla(\phi_1 - \phi_2)dv = \nabla^2(\phi_1 - \phi_2) = 0$$

and thus

$$\int_V (\phi_1 - \phi_2)\langle\nabla,\, \nabla(\phi_1 - \phi_2)\rangle dv = 0$$

Further, we can use the divergence theorem to replace the volume integral on the left by a surface integral which gives

$$\int_V \langle\nabla,\, (\phi_1 - \phi_2)\nabla(\phi_1 - \phi_2)\rangle dv = \oint_S \langle(\phi_1 - \phi_2)\nabla(\phi_1 - \phi_2),\, d\mathbf{S}\rangle = 0$$

and thus we have

$$\int_V \langle \nabla(\phi_1 - \phi_2),\, \nabla(\phi_1 - \phi_2) \rangle = 0$$

In this case, $\langle \nabla(\phi_1 - \phi_2),\, \nabla(\phi_1 - \phi_2) \rangle$ is everywhere nonnegative, thus the definite integral above can only be zero if

$$\langle \nabla(\phi_1 - \phi_2),\, \nabla(\phi_1 - \phi_2) \rangle = 0$$

and thus

$$\nabla(\phi_1 - \phi_2) = 0$$

which implies

$$\phi_1 - \phi_2 = \text{const.}$$

If any boundary condition is in place for the solution, such as $\lim_{x\to\infty} \phi(x) = 0$, then we have

$$\phi_1 - \phi_2$$

Note that in physically realistic scenarios we usually have at least one boundary condition.

8.4 Discretizing the Laplace operator

The discrete solution of this equation requires the discretization of the Laplace operator ∇^2. In a d-dimensional regular lattice with grid spacing h, the Laplace operator can be discretized as:

$$\nabla_d^2 f_i = \frac{1}{h^d} \left[\sum_{neighbors\ j} f_j - f_i \right]$$

For example, in 1D:

$$\nabla_1^2 f_i = \frac{f_{i-1} + f_{i+1} - 2f_i}{h}$$

In 2D:

$$\nabla_2^2 f_{i,j} = \frac{f_{i-1,j} + f_{i+1,j} + f_{i,j-1} + f_{i,j+1} - 4f_{i,j}}{h^2}$$

and in 3D:

$$\nabla_2^2 f_{i,j} = \frac{f_{i-1,j,k} + f_{i+1,j,k} + f_{i,j-1,k} + f_{i,j+1,k} + f_{i,j,k-1} + f_{i,j,k+1} - 6f_{i,j}}{h^3}$$

Using the 2D-case, we obtain the discretized Poisson equation:

$$\phi_{i-1,j} + \phi_{i+1,j} + \phi_{i,j-1} + \phi_{i,j+1} - 4\phi_{i,j} = \frac{h^2}{\epsilon_0} \rho_{i,j}.$$

This equation can be written in Matrix-vector form

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{x}$ contains all $\phi_{i,j}$, $\mathbf{b}$ contains the scaled potentials $h^2\rho_{i,j}$ and $\mathbf{A}$ contains the coefficients specifying the update rule. This can be solved directly.

8.5 Example: 1D

For example, let us consider the 1D case with the $n + 2$ grid points.

$$\phi_0 \quad \phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n \quad \phi_{n+1}$$

Now we use Dirichlet boundary conditions, i.e. we set the potential to the pre-defined values ϕ_0 and ϕ_{n+1} on the boundaries. For all intermediate points, the equation

$$\phi_{i-1} + \phi_{i+1} - 2\phi_i = -h\frac{\rho}{\epsilon_0}$$

must hold. Defining the matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and vector $\mathbf{b} \in \mathbb{R}^n$ on the points $1...n$,

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}$$

$$b_i = \begin{cases} -\frac{h}{\epsilon_0}\rho_i - \phi_0 & i = 1 \\ -\frac{h}{\epsilon_0}\rho_i - \phi_{n+1} & i = n \\ -\frac{h}{\epsilon_0}\rho_i & \text{else} \end{cases}$$

we solve $\mathbf{Ax} = \mathbf{b}$.

An alternative would be to use periodic boundary conditions. Consider the following grid

$$\phi_n \quad \phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_{n-1} \quad \phi_n$$

where $\phi_n = \phi_n$. The periodicity can be expressed by the shape of the matrix, as follows:

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & & & 1 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{bmatrix}$$

and we solve the system.

8.6 Example: 2D

Suppose we have the following field grid:

$$\begin{matrix} \phi_{0,0} & \phi_{0,1} & \cdots & \phi_{0,n} & \phi_{0,n+1} \\ \phi_{1,0} & \phi_{1,1} & \cdots & \phi_{1,n} & \phi_{0,n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ \phi_{n,0} & \phi_{n,1} & \cdots & \phi_{n,n} & \phi_{n,n+1} \\ \phi_{n+1,0} & \phi_{n+1,1} & \cdots & \phi_{n+1,n} & \phi_{n+1,n+1} \end{matrix}$$

where we again use Dirichlet boundary conditions. The inner window $\phi_{1,1}, ..., \phi_{n,n}$ are unknown potential values. Now we need to use a convention to vectorize this 2D grid. If we choose row-wise vectorization, i.e. we define the order

$$\phi_{1,1}, ..., \phi_{1,n}, \phi_{2,1}, ..., \phi_{2,n}, ..., \phi_{n,1}, ..., \phi_{n,n}$$

then we obtain the matrix ... (**try to figure the shape out yourself**).

8.7 Relaxation methods

The full solution of the linear system using Gauss elimination algorithm has a complexity of $\mathcal{O}(n^3)$, which is very unfavorable, especially if we are interested in computing energies and forces between N charges, which can be done in N^2 directly. Since the structure of the matrices above is very sparse, we can get down to $\mathcal{O}(n^2)$ using sparse matrix solvers.

Since, moreover the structure has a very specific neighborhood we can go further and write down relatively simple iterative sparse solvers. That means, we start with an initial potential ϕ^0 and then iteratively apply the discretized Poisson equation to each site in an update fashion. Common approaches include:

Jacobi method

Directly use the equation

$$\phi_{i,j}^{n+1} = \frac{\phi_{i-1,j}^{n} + \phi_{i+1,j}^{n} + \phi_{i,j-1}^{n} + \phi_{i,j+1}^{n} + h^2\rho_{i,j}}{4}$$

to perform an iteration that (hopefully) converges

Gauss-Seidel method

Sweep the lattice in order of increasing i and j

$$\phi_{i,j}^{n+1} = \frac{\phi_{i-1,j}^{n+1} + \phi_{i+1,j}^{n} + \phi_{i,j-1}^{n+1} + \phi_{i,j+1}^{n} + h^2\rho_{i,j}}{4}$$

Successive overrelaxation (SOR) method

Gauss-Seidel with linear combination of old and new solution:

$$\phi_{i,j}^{n+1} = (1-\omega)\phi_{i,j}^{n} + \omega\frac{\phi_{i-1,j}^{n+1} + \phi_{i+1,j}^{n} + \phi_{i,j-1}^{n+1} + \phi_{i,j+1}^{n} + h^2\rho_{i,j}}{4}$$

where the theoretical optimum for a 2D-lattice is $\omega = 2/(1+\pi/L)$, L being the number of lattice points in x/y direction.

Checkerboard updating

SOR with iterative updating of lattice sites that have:

$$\begin{array}{ll} i+j & \text{odd} \\ i+j & \text{even} \end{array}$$

Convergence rates

Splitting matrix $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{U}$ into a lower triangle, diagonal, upper triangle matrix, we can write the Matrix-vector form as:

$$\begin{aligned} (\mathbf{L}+\mathbf{D}+\mathbf{U})\mathbf{x} &= \mathbf{b} \\ \mathbf{D}\mathbf{x} &= \mathbf{b} - (\mathbf{L}+\mathbf{U})\mathbf{x} \end{aligned}$$

where the left-hand $\mathbf{x}$ can be identified with $\mathbf{x}^{n+1}$ and the right-hand $\mathbf{x}$ with $\mathbf{x}^{n}$ in the Jacobi iteration:

$$\mathbf{x}^{n+1} = -\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})\mathbf{x}^{n} + \mathbf{D}^{-1}\mathbf{b}.$$

Here the matrix

$$\mathbf{I} = -\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})$$

is the iteration matrix. Its spectral radius (largest eigenvalue) determines whether and how fast the iteration converges towards the solution. The spectral radius must lie between 0 and 1 to be stable, depends on the boundary conditions, and approaches 1 as the number of lattice sites increases.

For the methods above we find the following spectral radii, and accordingly the following number of steps n to reduce the error by a factor 10:

- Jacobi: $\rho_s \approx 1 - \frac{\pi^2}{2L^2} \rightarrow n \approx \frac{1}{2}L^2$
- Gauss-Seidel: $\rho_s \approx 1 - \frac{\pi^2}{L^2} \rightarrow n \approx \frac{1}{4}L^2$
- Successive Overrelaxation: $\rho_s \approx 1 - \frac{2\pi}{L} \rightarrow n \approx \frac{1}{3}L$

Thus, SOR can offer dramatic improvements over Jacobi and Gauss-Seidel.

8.8 Exercises

1. **Poisson Equation**: Write down the two-dimensional finite-difference discretization of the unitless Poisson equation ($\nabla^2\phi_i(\mathbf{r}) = -\rho_i(\mathbf{r})$) on a uniform grid with cell length d. Bring this equation into the form $\mathbf{A}\boldsymbol{\phi} = \mathbf{b}$ for the case of periodic boundary conditions. State $\mathbf{A}$ and $\mathbf{b}$.

2. **Poisson Solution**: Use the periodic 2D-System $[0, 10] \times [0, 10] \subset \mathbb{R}^2$ with charge $q_1 = +1$ on $\mathbf{r}_1 = (2.5, 2.5)$ and $q_2 = -1$ on $\mathbf{r}_2 = (7.5, 7.5)$. Solve the Poisson equation on a grid with size $d = 0.1$.

3. **Relaxation method**: Implement a solution of 3. using the Jacobi relaxation method. Show the error in ϕ ($\epsilon = \sqrt{\sum_{i,j}\left(\phi_{i,j}^{(1)} - \phi_{i,j}^{(2)}\right)^2}$) with respect to the Matrix solution as a function of the number of Jacobi iterations.

9 Discrete and Fast Fourier Transform

Disclaimer: Contains material from https://en.wikipedia.org/wiki/Fourier_ transform

The Fourier transform decomposes a real-valued signal (e.g. a function of time) into the frequencies that make it up, in a way similar to how a musical chord can be expressed as the amplitude (or loudness) of its constituent notes. Mathematically, a Fourier transform is an invertible mapping of the real-space function $x(t)$ with $t \in \mathbb{R}$ to its complex-valued transform $X(\omega)$ with $\omega \in \mathbb{C}$:

$$x(t) \quad \underset{\text{inverse Fourier transform}}{\overset{\text{Fourier transform}}{\rightleftarrows}} \quad X(\omega)$$

Intuitively, $x(t)$ contains the value of the signal at any point in time, whereas $X(\omega)$ contains, for each frequency or mode, the amplitude and the phase (shift) which which it is contained in the signal. In general, t may not correspond to time but some other physical variable (e.g. space), or it may be a purely mathematical object with no direct physical interpretation. Moreover, the Fourier transform can be done in multiple dimensions - we will present the one-dimensional Fourier transform first, and then state the multidimensional form.

Fourier transforms are widely used in signal analyses. In particular the discrete Fourier transform (DFT) and the fast Fourier transform (FFT, which is equivalent to the DFT but uses a more efficient algorithm) is a key technology for time-discrete or time-discretized data that are ubiquitous in digital signals. Applications of DFT and FFT include:

- Efficient approximation of continuous-time Fourier transforms
- Efficient computation of convolutions using the relation $x(t) * v(t) \leftrightarrow X(\omega)V(\omega)$, provided $x(t)$ and $v(t)$ are zero outside some common range $t = 0, ..., T-1$.
- Removal of noise from sampled data - the idea is that "white noise" has a very special frequency-domain behavior, which makes it easy to detect and remove.

Notation:

T timesteps, time index t

Frequency ω

Signal $\{x_t\} = (x_0, ..., x_{T-1})$

Fourier transformed signal $\{X_k\} = (X_0, ..., X_{T-1})$

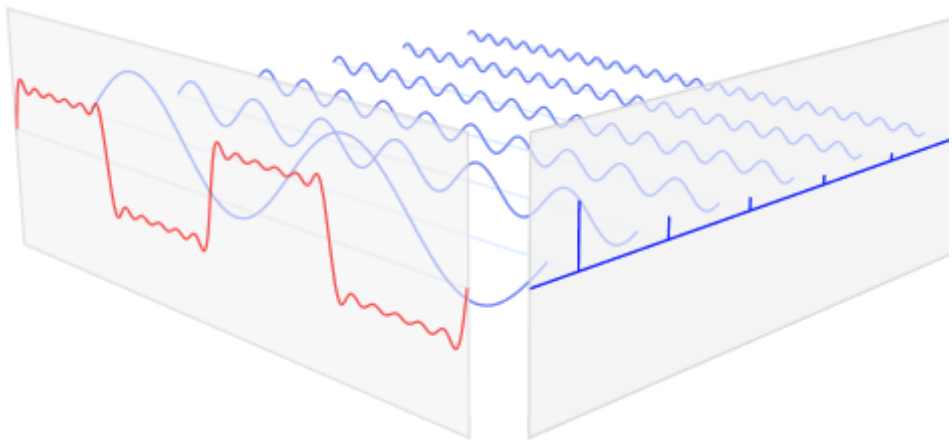
Note: not a good notation if we need matrices later.

Complex number representation:

$$re^{-i\omega} = r\left(\cos(\omega) + i\sin(\omega)\right).$$

9.1 Fourier series and Fourier transform

We will first start with the **Fourier series**, which is a special case of the Fourier transform but they are more intuitive for the beginning. Fourier series can be applied to **periodic** signals. Consider the following decomposition:



It can be seen that the sum of sine functions with higher and higher frequencies (blue) can add up to a periodic function of different shape, here a step function. If we keep adding up more higher-frequency functions we will eventually converge to a perfect step function, but infinitely many frequencies are needed in order to form a perfect step. We can see that the real-space representation of the signal (red) is one amplitude per time, $x(t)$. To make up the Fourier representation of the signal we need to store an amplitude and a phase shift for each frequency occurring in the signal. For every frequency, we thus need two real numbers, and we do that by using a complex number

$$re^{-i\omega} = r\left(\cos\omega + i\sin\omega\right),$$

where r is the amplitude and ω is the phase. The Fourier series of a periodic signal $x(t)$ is then given by

$$X(\omega) = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\mathrm{e}^{-i\omega t}\,\mathrm{d}t$$

where we have the discrete frequencies

$$\omega = \frac{2\pi n}{T}$$

where n is the number of oscillations per period T. The inverse Fourier series, i.e. the reconstruction of the signal using the Fourier representation is given by

$$x(t) = \sum_{\omega=-\infty}^{\infty} X(\omega)\mathrm{e}^{i\omega t}.$$

If $x(t)$ is periodic in T, this mapping reverses the Fourier transform, i.e. we have a bijective mapping between $x(t)$ and $X(\omega)$.

The Fourier transform is the generalization of the Fourier transform to arbitrary (nonperiodic) signals. The trick is to let $T \to \infty$, i.e. to allow every event in the signal to only occur once during the entire observation range. In this limit, the spacing between discrete frequencies ω shrinks and we get a continuous spectrum. The Fourier transform is defined quite straightforwardly as

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-i\omega t}\,\mathrm{d}t$$

and the inverse Fourier transform is again

$$x(t) = \sum_{\omega=-\infty}^{\infty} X(\omega)\mathrm{e}^{i\omega t}.$$

And we introduce the notation:

$$X(\omega) = \mathcal{F}\left\{x(t)\right\}$$
$$x(t) = \mathcal{F}^{-1}\left\{X(\omega)\right\}.$$

Now, the invertibility $x(t) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{x(t)\right\}\right\}$ holds for any real-valued signal.

9.2 Properties and Applications

The Fourier transform has the following basic properties:

1. **Linearity**: For any complex numbers a and b, $\mathcal{F}\{ax(t) + by(t)\} = a\mathcal{F}\{x(t)\} + b\mathcal{F}\{y(t)\}$.
2. **Translation/Time-Shifting**: For any real number τ, $\mathcal{F}\{x(t-\tau)\} = \mathrm{e}^{-i\omega\tau}\mathcal{F}\{x(t)\}$.
3. **Modulation/Frequency shifting**: For any real number ω_0, $\mathcal{F}\left\{\mathrm{e}^{i\omega_0 t}x(t)\right\}(\omega) = \mathcal{F}\{x(t)\}(\omega - \omega_0)$.
4. **Time Scaling**: For a non-zero real number a, $\mathcal{F}\{x(at)\} = \frac{1}{|a|}\mathcal{F}\{x(t)\}(\frac{\omega}{a})$. The case $a = -1$ leads to the time-reversal property, which states: if $\mathcal{F}\{x(-t)\}(\omega) = \mathcal{F}\{x(t)\}(-\omega)$.
5. **Integration**: Substituting $\omega = 0$ in the definition, we obtain $X(0) = \int_{-\infty}^{\infty} x(t)\,\mathrm{d}t$, i.e. the zero frequency equals the integral over the whole domain.

and following theorems can be derived:

1. **Differentiation**: $\mathcal{F}\left\{\frac{\mathrm{d}^n}{(\mathrm{d}t)^n}x(t)\right\} = (i\omega)^n\mathcal{F}\{x(t)\}$. By applying the Fourier transform and using these formulas, some ordinary differential equations can be transformed into algebraic equations, which are much easier to solve.
2. **Convolution**: If
$$z(\tau) = (x * y)(\tau) = \int_{-\infty}^{\infty} x(t)\,y(\tau - t)\,\mathrm{d}\tau$$
Then
$$\mathcal{F}\{x * y\} = \mathcal{F}\{x\}\,\mathcal{F}\{y\}$$
3. **Cross-correlation**: If
$$z(\tau) = \mathrm{corr}(x,\, y) = \int_{-\infty}^{\infty} x(t)\,y(t + \tau)\,\mathrm{d}t$$
Then
$$\mathcal{F}\{\mathrm{corr}(x,\, y)\} = \mathcal{F}\{x\}\,\mathcal{F}\{y\}$$

9.3 Discrete Fourier Transform

Now we consider the situation that our signal is discrete. This may happen because a continuous signal has been sampled at some sampling interval, or the signal is discrete by nature (i.e. a vector). In order to keep the same symbols as above, we slightly modify their meaning, and define the signal vector:

$$\mathbf{x} = [x_n] = (x_0,\, ...,\, x_{N-1})$$

where N corresponds to the number of sampled time steps or spatial positions. Then the N-point Discrete Fourier Transform (DFT) of x_n is defined as a vector

$$\mathbf{X} = [X_k] = (X_0,\, ...,\, X_{N-1}),$$

where

$$X_k = \sum_{n=0}^{N-1} x_n \mathrm{e}^{-i2\pi\frac{nk}{N}}.$$

The reverse transform (inverse DFT) is defined by:

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \mathrm{e}^{i2\pi\frac{nk}{N}}$$

Proof:

$$\begin{aligned} x_n &= \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{e}^{i2\pi\frac{nk}{N}} \sum_{n'=0}^{N-1} x_{t'}\mathrm{e}^{-i2\pi\frac{n'k}{N}} \\ &= \frac{1}{N}\sum_{n'=0}^{N-1} x_{n'} \sum_{k=0}^{N-1} \mathrm{e}^{i2\pi\frac{k}{N}(n-n')} \end{aligned}$$

The inner sum evaluates to:

$$\sum_{k=0}^{N-1} \mathrm{e}^{i2\pi\frac{k}{N}(n-n')} = \begin{cases} N & n = n' \\ 0 & n \neq n' \end{cases}$$

which proofs the reconstruction $\mathbf{x} = \mathrm{DFT}^{-1}(\mathrm{DFT}(\mathbf{x}))$.

Computational complexity: Direct evaluation of both forward and backward transform involve T multiplications for each of T time steps or frequencies, i.e. the direct DFT has complexity $\mathcal{O}(T^2)$.

9.4 Fast Fourier Transform

Discovered by Gauss (around 1805), but usually attributed to Cooley and Tuckey (1965). We assume that $N = 2^n$ with some integer n, i.e. we have a power of 2 time indexes. If not, then we can just pad the original signal to achieve a length that is a power of 2. Let us call

$$W_N = \mathrm{e}^{-i2\pi\frac{1}{N}}$$

which is a Nth root of unity, i.e.

$$(W_N)^N = \mathrm{e}^{-i2\pi} = 1.$$

We can write the DFT as

$$X_k = \sum_{n=0}^{N-1} x_n (W_N)^{nk}$$

Now we divide the signal $\mathbf{x} = (x_0, ..., x_{N-1})$ into two subsignals $\mathbf{a}$ and $\mathbf{b}$ of lengths $N/2$ as:

$$\begin{aligned} a_n &= x_{2n} \qquad n = 0, 1, ..., \frac{N}{2} - 1 \\ b_n &= x_{2n+1} \quad n = 0, 1, ..., \frac{N}{2} - 1 \end{aligned}$$

Thus, $\mathbf{a}$ contains the even and $\mathbf{b}$ contains the odd time-indexes of the original signal. We compute their DFT's as:

$$\begin{aligned} A_k &= \sum_{n=0}^{\frac{N}{2}-1} a_n \mathrm{e}^{-i2\pi\frac{nk}{N/2}} = \sum_{n=0}^{\frac{N}{2}-1} a_n (W_{N/2})^{nk} \\ B_k &= \sum_{n=0}^{\frac{N}{2}-1} b_n \mathrm{e}^{-i2\pi\frac{nk}{N/2}} = \sum_{n=0}^{\frac{N}{2}-1} b_n (W_{N/2})^{nk} \end{aligned}$$

The main idea of the FFT is that we can reconstruct the DFT of $\mathbf{x}$ by the DFTs of $\mathbf{a}$ and $\mathbf{b}$ by:

$$X_k = \begin{cases} A_k + (W_N)^k B_k & k = 0, ..., \frac{N}{2} - 1 \\ A_{k-\frac{N}{2}} - (W_N)^k B_k & k = \frac{N}{2}, ..., N - 1 \end{cases}$$

Note that this procedure can be recursively applied to $\mathbf{a}$ and $\mathbf{b}$ until we only have a single number (x_0) left, for which the transform is just $X_0 = x_0$. For each value of k, we have on the order of $\log_2$ multiplications and additions, overall the complexity of the FFT is $\mathcal{O}(N \log_2 N)$ rather than $\mathcal{O}(N^2)$.

Proof: The first part is proven by

$$A_k + (W_N)^k B_k = \sum_{n=0}^{\frac{N}{2}-1} a_n (W_{N/2})^{nk} + (W_N)^k \sum_{n=0}^{\frac{N}{2}-1} b_n (W_{N/2})^{nk}$$

Note that

$$(W_{N/2})^{nk} = \mathrm{e}^{-i2\pi \frac{nk}{N/2}} = \mathrm{e}^{-i2\pi \frac{2nk}{N}} = (W_N)^{2nk}.$$

And thus:

$$\begin{aligned} A_k + (W_N)^k B_k &= \sum_{n=0}^{\frac{N}{2}-1} a_n (W_N)^{2nk} + \sum_{n=0}^{\frac{N}{2}-1} b_n (W_N)^{(2n+1)k} \\ &= \sum_{n=0}^{\frac{N}{2}-1} x_{2n} (W_N)^{2nk} + \sum_{n=0}^{\frac{N}{2}-1} x_{2n+1} (W_N)^{(2n+1)k} \\ &= \sum_{n=0}^{N-1} x_n (W_N)^{nk} = X_k. \end{aligned}$$

The second part works analogously, and the same principles can be derived the inverse FFT.

9.5 Multidimensional DFT / FFT

In order to solve the Poisson equation using the DFT, we will not do a Fourier transform of time, but rather of the space. However, we have three spatial coordinates. Let us in general define the d-dimensional vectors

$$\begin{aligned} \mathbf{n} &= (n_1, \ldots, n_d)^\top \\ \mathbf{k} &= (k_1, \ldots, k_d)^\top \end{aligned}$$

where the indexes $\mathbf{n}$ range from $\mathbf{0}$ to $\mathbf{N} = (N_1, \ldots, N_d)$. Then, the multidimensional DFT of $x_{\mathbf{n}}$ is:

$$X_{\mathbf{k}} = \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N}-1} e^{-2\pi i \mathbf{k} \cdot (\mathbf{n}/\mathbf{N})} x_{\mathbf{n}}$$

where the division $\mathbf{n}/\mathbf{N}$ is defined as $\mathbf{n}/\mathbf{N} = (n_1/N_1, \ldots, n_d/N_d)$ to be performed element-wise. The inverse of the multi-dimensional DFT is, analogous to the one-dimensional case, given by:

$$x_{\mathbf{n}} = \frac{1}{\prod_{\ell=1}^{d} N_\ell} \sum_{\mathbf{k}=\mathbf{0}}^{\mathbf{N}-1} e^{2\pi i \mathbf{n} \cdot (\mathbf{k}/\mathbf{N})} X_{\mathbf{k}}$$

As the one-dimensional DFT expresses the input x_{n} as a superposition of sinusoids, the multidimensional DFT expresses the input as a superposition of plane waves, or multidimensional sinusoids. The direction of oscillation in space is $\mathbf{k}/\mathbf{N}$. The amplitudes are $X_{\mathbf{k}}$.

The multidimensional DFT can be computed by the composition of a sequence of one-dimensional DFTs along each dimension. In the two-dimensional case x_{n_1,n_2} the N_1 independent DFTs of the rows (i.e., along n_2) are computed first to form a new array y_{n_1,k_2}. Then the N_2 independent DFTs of y along the columns (along n_1) are computed to form the final result X_{k_1,k_2}. Alternatively the columns can be computed first and then the rows. The order is irrelevant because the nested summations above commute. Thus, an algorithm to compute a one-dimensional DFT efficiently is thus sufficient to compute a multidimensional DFT efficiently. We use the FFT algorithm described above for that purpose.

9.6 Solving the Poisson equation with DFT/FFT

As a reminder: We want to solve the Poisson equation

$$\nabla^2\phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0}$$

with charge density $\rho(\mathbf{r})$ and electric potential $\phi(\mathbf{r})$ on a grid with space indexes $\mathbf{r} = (x,\, y,\, z)^\top$. Since we consider a discretization on a regular lattice, the mesh coordinates can be given by:

$$\mathbf{r_m} = \frac{n_1}{N_1}\mathbf{a}_1 + \frac{n_2}{N_2}\mathbf{a}_2 + \frac{n_3}{N_3}\mathbf{a}_3$$

where we have the index set

$$\mathbf{n} = (n_1,\, n_2,\, n_3)$$

and the box extensions

$$\mathbf{a} = (a_1,\, a_2,\, a_3)$$

The Laplace operator is given by $\nabla^2 = \frac{\mathrm{d}^2}{\mathrm{d}x} + \frac{\mathrm{d}^2}{\mathrm{d}y} + \frac{\mathrm{d}^2}{\mathrm{d}z}$. Using the linearity and differentiation properties of the Fourier transform, we get

$$\begin{aligned}
\nabla^2\phi(\mathbf{r}) &= \mathcal{F}\left\{\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{\mathrm{d}^2}{\mathrm{d}y^2} + \frac{\mathrm{d}^2}{\mathrm{d}z^2}\right)\phi(\mathbf{r})\right\} \\
&= \mathcal{F}\left\{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\phi(\mathbf{r})\right\} + \mathcal{F}\left\{\frac{\mathrm{d}^2}{\mathrm{d}y^2}\phi(\mathbf{r})\right\} + \mathcal{F}\left\{\frac{\mathrm{d}^2}{\mathrm{d}z^2}\phi(\mathbf{r})\right\} \\
&= \left((ik_x)^2 + (ik_y)^2 + (ik_z)^2\right)\mathcal{F}\left\{\phi(\mathbf{r})\right\} \\
&= -k^2\mathcal{F}\left\{\phi(\mathbf{r})\right\}
\end{aligned}$$

with $k = k_x^2 + k_y^2 + k_z^2$. Furthermore, using linearity we have

$$\mathcal{F}\left\{-\frac{\rho(\mathbf{r})}{\epsilon_0}\right\} = -\frac{1}{\epsilon_0}\mathcal{F}\left\{\rho(\mathbf{r})\right\}$$

and thus the Fourier-space Poisson equation is

$$k^2\Phi(\mathbf{k}) = \frac{1}{\epsilon_0}P(\mathbf{k})$$

with $\Phi(\mathbf{k}) = \mathcal{F}\left\{\phi(\mathbf{r})\right\}$ and $P(\mathbf{k}) = \mathcal{F}\left\{\rho(\mathbf{r})\right\}$ and the finite Fourier space lattice:

$$\mathbf{k_n} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2 + n_3\mathbf{b}_3$$

Therefore we can solve for ϕ by:

1. Transform charge density to Fourier space: $P(\mathbf{k}) = \mathcal{F}\{\rho(\mathbf{r})\}$
2. Compute Fourier-space potential: $\Phi(\mathbf{k}) = \frac{P(\mathbf{k})}{k^2\epsilon_0}$.
3. Transform back to real space: $\phi(\mathbf{r}) = \mathcal{F}^{-1}\{\Phi(\mathbf{k})\}$

The discrete Fourier transforms can be performed using the FFT. Assuming that we assign m lattice points to each of N charges, this results in a complexity of $Nm\log Nm = m(N\log N + N\log m)$ which has formal complexity $N\log N$ with a possibly large pre-factor m.

10 Electrostatics in periodic systems - Ewald Summation

10.1 Convergence in periodic systems

The total potential energy at position $\mathbf{r}_i$ is given by

$$V_s(\mathbf{r}_i) = \sum_{\mathbf{n}}{}' \sum_{j=1}^{N} \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \tag{9}$$

where the sum $\sum_{\mathbf{n}}'$ runs over all periodic box vectors, and the prime indicates that the term $i = j$ is omitted in the primary cell. The vector $\mathbf{n} = (i_x, i_y, i_z)L$ points to the origin of a periodic box, with indexes $i_{x,y,z} = 0, \pm 1, \pm 2, \ldots$. The sum (9) is only conditionally convergent. For example take the following one-dimensional charge series:

$$\cdots + - + - + - \cdots$$

where charges are separated by a fixed distance d. Consider two ways of computing the electrostatic interaction energy of the central two $(+-)$ charges:

1. Consider the periodic system:
$$\cdots (+-)(+-)(+-) \cdots$$
and first compute the interaction energy of the primary cell, then add all interaction energies of the primary with the secondary cells, then add all interaction energies of the primary with the teriary cell, etc. Iterate this over periodic image pairs until convergence.

2. Consider this iteration:
$$\cdots (+ (- (+) -) +) \; - \; \cdots$$
i.e. add up the interaction energy of the central + with its nearest-neighbor − charges. Then add the interaction energy with the next + charges, etc. Continue adding neighbors until convergence. For symmetry reasons, the interaction energy of the central − with the rest:
$$\cdots + \; (- (+ (-) +) -) \cdots$$
will be equal, so we just double the result.

It turns out that (1) will not converge, but (2) will. In general, taking image cells in the order of increasing distance from the primary cell ensures that V_s converges. However, even then the convergence is very very slow.

10.2 Ewald summation

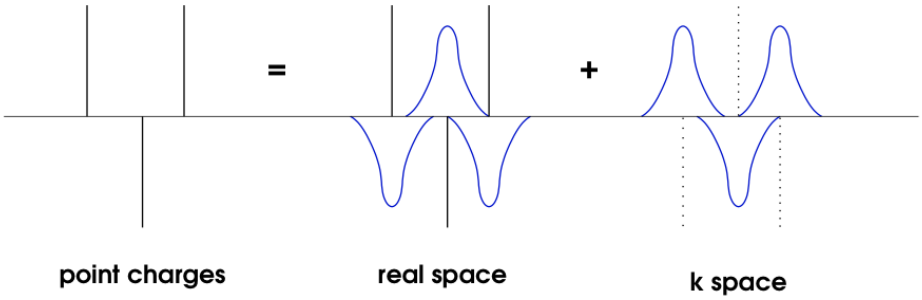
Idea: We recompute the electrostatic energy, but rewrite the charge distribution by subtracting and adding subtracting a smeared charge distribution $G_\sigma(\mathbf{r})$:

$$\rho_i(\mathbf{r}) \quad = \quad \underbrace{q_i\delta(\mathbf{r} - \mathbf{r}_i) - q_i G_\sigma(\mathbf{r} - \mathbf{r}_i)}_{\rho_i^S(\mathbf{r})} + \underbrace{q_i G_\sigma(\mathbf{r} - \mathbf{r}_i)}_{\rho_i^L(\mathbf{r})}$$

and split this sum into parts $\rho_i^S(\mathbf{r})$ and $\rho_i^L(\mathbf{r})$. This approach is useful because with an appropriate choice of G_σ the contribution $-q_i G_\sigma(\mathbf{r} - \mathbf{r}_i)$ counteracts (shields) the charge q_i such that the part $\rho_i^S(\mathbf{r})$ only generates a short-ranged field that decays quickly in real space can be efficiently truncated using a real-space cutoff. We are left with $\rho_i^L(\mathbf{r})$ which does generate a long-ranged potential, but, is periodic on the lattice and with an appropriate choice of G_σ is smooth enough such that we can efficiently describe it as a Fourier sum using not too many Fourier coefficients. For G_σ various choices are possible, but we will use a normalized Gaussian distribution:

$$G_\sigma(\mathbf{r}) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

where $r = |\mathbf{r}|$. The scheme below illustrates the charge splitting of a set of point charges into shielded point charges to be evaluated as a direct real space sum and shielding potentials that will be transformed to reciprocal space, solved there, and then back-transformed to real space.



Potential decomposition:

$$\phi_i(\mathbf{r}) = \underbrace{\frac{q_i}{4\pi\epsilon_0}\left[\frac{1}{|\mathbf{r}_i-\mathbf{r}|}-\int d^3\mathbf{r}'\frac{G_\sigma(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\right]}_{\phi_i^S(\mathbf{r})} + \underbrace{\frac{q_i}{4\pi\epsilon_0}\int d^3\mathbf{r}'\frac{G_\sigma(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}}_{\phi_i^L(\mathbf{r})} \tag{10}$$

Short range potential and interaction energy

The potential field generated by a Gaussian charge distribution is evaluated by solving Poisson's equation $\nabla^2\phi_\sigma(\mathbf{r}) = -G_\sigma(\mathbf{r})/\epsilon_0$. This has been done above and the result is given in Eq. (5). Using this result and associating $r = |\mathbf{r}-\mathbf{r}_i|$ for each charge, we can thus obtain the electric potential $\phi_i^S(\mathbf{r})$ of the decomposition in Eq. (10):

$$\phi_i^S(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0|\mathbf{r}_i-\mathbf{r}|}\left[1-\mathrm{erf}\left(\frac{|\mathbf{r}_i-\mathbf{r}|}{\sqrt{2}\sigma}\right)\right] = \frac{q_i}{4\pi\epsilon_0|\mathbf{r}_i-\mathbf{r}|}\mathrm{erfc}\left(\frac{|\mathbf{r}_i-\mathbf{r}|}{\sqrt{2}\sigma}\right) \tag{11}$$

The error function erf quickly increases from $\mathrm{erf}(0) = 0$ to a value of 1, while the complementary error function erfc quickly decays from $\mathrm{erfc}(0) = 1$ to 0. Therefore, can associate the shielded potential $\phi_i^S(\mathbf{r})$ is a short-range potential that can be efficiently cut off. We therefore compute the corresponding electrostatic energy directly between pairs of charges, and will later truncate this term at a certain cutoff radius:

$$E^S = \frac{1}{8\pi\epsilon_0}\sum_{\mathbf{n}}\sum_{i=1}^{N}{}^*\frac{q_iq_j}{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}L|}\mathrm{erfc}\left(\frac{|\mathbf{r}_i-\mathbf{r}_j+\mathbf{n}L|}{\sqrt{2}\sigma}\right) \tag{12}$$

Here the star $\sum^*$ indicates that the term $i = j$ is omitted in the primary cell in order to avoid interaction of the charge with itself.

Long range potential and interaction energy

The potential $\phi_i^L(\mathbf{r})$ is dominated by a r^{-1} decay, and thus a long-range potential that cannot be efficiently cutoff:

$$\phi_i^L(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0|\mathbf{r}_i-\mathbf{r}|}\mathrm{erf}\left(\frac{|\mathbf{r}_i-\mathbf{r}|}{\sqrt{2}\sigma}\right). \tag{13}$$

The total long-ranged charge density

$$\rho^L(\mathbf{r}) = \sum_{\mathbf{n}}\sum_{j=1}^{N} q_jG_\sigma(\mathbf{r}-\mathbf{r}_j+\mathbf{n}L) \tag{14}$$

is periodic in $\mathbf{n}$, and likewise the total potential ϕ^L is periodic. Therfore we consider to solve it by Fourier transforming the charge distribution to the reciprocal space, solving for ϕ^L there and then transforming it back. We will see that in this way we will bet an expression for the potential and total energy that can also be efficiently cut off, only that now the cut-off parameter is not a real-space radius but rather a reciprocal space radius, i.e. we restrict ourselves to considering only a few wave vectors.

Consider the Fourier transforms of ρ^L and ϕ^L, denoted by $\hat{\rho}^L = \mathcal{F}\{\rho^L(\mathbf{r})\}$ and $\hat{\phi}^L = \mathcal{F}\{\phi^L(\mathbf{r})\}$. The forward and backward Fourier transforms are given by:

$$\begin{array}{ll} \hat{\phi}^L(\mathbf{k}) = \int_V d^3\mathbf{r}\, \phi^L(\mathbf{r}) \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}\rangle} & \phi^L(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}} \hat{\phi}^L(\mathbf{k}) \mathrm{e}^{i\langle \mathbf{k},\mathbf{r}\rangle} \\ \hat{\rho}^L(\mathbf{k}) = \int_V d^3\mathbf{r}\, \rho^L(\mathbf{r}) \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}\rangle} & \rho^L(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}} \hat{\rho}^L(\mathbf{k}) \mathrm{e}^{i\langle \mathbf{k},\mathbf{r}\rangle} \end{array}$$

where $\mathbf{k} = \frac{2\pi}{L}(k_1, k_2, k_3)$ are reciprocal lattice vectors, where $k_{1,2,3} \in \mathbb{Z}$.

We Fourier-transform the charge density (14). Due to periodicity we simplify the calculation by considering only the charges of the primary cell and take the integral over the entire $\mathbb{R}^3$:

$$\begin{aligned} \hat{\rho}^L(\mathbf{k}) &= \sum_{j=1}^{N} q_j \int_{\mathbb{R}^3} d^3\mathbf{r}\, G_\sigma(\mathbf{r} - \mathbf{r}_j) \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}\rangle} \\ &= \sum_{j=1}^{N} q_j \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}_j\rangle} \mathrm{e}^{-\sigma^2 k^2/2} \end{aligned}$$

where $k = |\mathbf{k}|$ is the reciprocal vector length. Now we compute the reciprocal potential using the reciprocal Poisson equation (**??**):

$$\hat{\phi}^L(\mathbf{k}) = \frac{1}{\epsilon_0} \sum_{j=1}^{N} q_j \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}_j\rangle} \frac{\mathrm{e}^{-\sigma^2 k^2/2}}{k^2}$$

and compute the real-space potential by back-transforming $\hat{\phi}^L(\mathbf{k})$:

$$\phi^L(\mathbf{r}) = \frac{1}{V\epsilon_0} \sum_{\mathbf{k}} \sum_{j=1}^{N} q_j \mathrm{e}^{i\langle \mathbf{k},\mathbf{r}\rangle} \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}_j\rangle} \frac{\mathrm{e}^{-\sigma^2 k^2/2}}{k^2}$$

Note that the contribution from the $\mathbf{k} = \mathbf{0}$ term is zero if the total charge is zero, i.e. $\sum_i q_i = 0$. We now compute the total long-range interaction energy, using indexes a, b in order to avoid confusions with the imaginary unit i:

$$\begin{aligned} E^L &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k}} \sum_{a=1}^{N} \sum_{b=1}^{N} q_a q_b \mathrm{e}^{i\langle \mathbf{k},\mathbf{r}_a\rangle} \mathrm{e}^{-i\langle \mathbf{k},\mathbf{r}_b\rangle} \frac{\mathrm{e}^{-\sigma^2 k^2/2}}{k^2} \\ &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k}} |S(\mathbf{k})|^2 \frac{\mathrm{e}^{-\sigma^2 k^2/2}}{k^2} \end{aligned} \tag{15}$$

where we have defined the structure factor

Using the structure factor, the calculation of the long-ranged energy is linear in N. The sum over $\mathbf{k}$ is infinite, but can be efficiently truncated by choosing a maximal cutoff for k, because the term $\mathrm{e}^{-\sigma^2 k^2/2}/k^2$ decays quickly in k.

Self-interaction energy

The long-ranged interaction energy is given by the sum of the interaction energies of each charge with the potential generated by the other charges. However, we have so far ignored that a given charge should not interact with itself. In the short-range energy 12 this is easy to take care of, by simply excluding the pair $i = j$ for the principal cell. However, in the long-range energy we have avoided such an "exception" for the primary cell so as to make sure that the lattice is entirely periodic. Therefore, this is not yet taken care of: the Gaussian charge centered at position $\mathbf{r}_i$ does generate a nonzero potential at $\mathbf{r}_i$, and hence E_L includes a nonphysical self-interaction that needs to be subtracted. This spurious self-interaction is easy to compute by considering the long-range potential given by Eq. (13) and letting $\mathbf{r} \to \mathbf{r}_i$:

$$\lim_{\mathbf{r}\to\mathbf{r}_i} \phi_i^L(\mathbf{r}) = \lim_{\mathbf{r}\to\mathbf{r}_i} \frac{q_i}{4\pi\epsilon_0|\mathbf{r}_i - \mathbf{r}|} \mathrm{erf}\left(\frac{|\mathbf{r}_i - \mathbf{r}|}{\sqrt{2}\sigma}\right) = \frac{q_i}{4\pi\epsilon_0\sigma}\sqrt{\frac{2}{\pi}},$$

yielding a total self-energy generated by all charges of:

$$E^{\text{self}} = \frac{1}{2\epsilon_0\sigma(2\pi)^{3/2}}\sum_{i=1}^{N} q_i^2 \tag{16}$$

Total Ewald interaction energy

The total Ewald interaction energy is thus given by combining Eqs. (12), (15) and (16):

$$\begin{aligned} E &= E^S + E^L - E^{\text{self}} \\ &= \frac{1}{8\pi\epsilon_0}\sum_{\mathbf{n}}\sum_{i=1}^{N}\sum_{i=1}^{N}{}^{*}\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|}\text{erfc}\left(\frac{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|}{\sqrt{2}\sigma}\right) \\ &\quad + \frac{1}{2V\epsilon_0}\sum_{\mathbf{k}}|S(\mathbf{k})|^2\frac{\mathrm{e}^{-\sigma^2k^2/2}}{k^2} \\ &\quad - \frac{1}{2\epsilon_0\sigma(2\pi)^{3/2}}\sum_{i=1}^{N} q_i^2 \end{aligned} \tag{17}$$

10.3 Interaction forces

The forces are given by the negative gradient:

$$\mathbf{f}(\mathbf{r}) = -\nabla E(\mathbf{x})$$

Let us differentiate the individual terms:

$$\begin{aligned} \mathbf{f}_i^S &= \frac{q_i}{8\pi\epsilon_0}\sum_{\mathbf{n}}\sum_{j=1}^{N}{}^{*}q_j\frac{(\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|^2}\left[\frac{\text{erfc}\left(\frac{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|}{\sqrt{2}\sigma}\right)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|} + \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}L|^2}{2\sigma^2}\right)\right] \\ \mathbf{f}_i^L &= \frac{1}{2V\epsilon_0}\sum_{\mathbf{k}}\frac{\mathrm{e}^{-\sigma^2k^2/2}}{k^2}2S(\mathbf{k})\frac{d}{d\mathbf{r}_i}S(\mathbf{k}) \\ \frac{d}{d\mathbf{r}_i}S(\mathbf{k}) &= q_i\mathrm{e}^{i\langle\mathbf{k},\mathbf{r}_i\rangle}\mathbf{k} \\ \mathbf{f}_i^L &= \frac{q_i}{V\epsilon_0}\sum_{\mathbf{k}}\frac{\mathrm{e}^{-\sigma^2k^2/2}}{k^2}S(\mathbf{k})\mathrm{e}^{i\langle\mathbf{k},\mathbf{r}_i\rangle}\mathbf{k} \end{aligned}$$

10.4 Approximation and approximation error

In order to compute Eq. (17) efficiently, we must truncate the infinite sums. We would like to control the error made by this, both in the real-space and in the reciprocal space sum.

First we set cutoff Radius R such that the cutoff error is given by e^{-p}:

$$\text{erfc}(\frac{r}{\sqrt{2}\sigma})\ |_{r=R} \approx \mathrm{e}^{-R^2/2\sigma^2} = \mathrm{e}^{-p}$$

This yieldsyielding a constraint on σ:

$$\sigma = \frac{R}{\sqrt{2p}} \tag{18}$$

Next, the k-space error can be parametrized as:

$$\begin{aligned} \mathrm{e}^{-\sigma^2K^2/2} &= \mathrm{e}^{-p} \\ p &= \frac{\sigma^2K^2}{2} \end{aligned} \tag{19}$$

where K is the k-space cutoff. Combining constraints (18) and (19) yields the relation:

$$K = \frac{\sqrt{2p}}{\sigma} = \frac{2p}{R} \tag{20}$$

Once the accuracy (via p) and the cutoff radius R have been chosen, Equations (18) and (20) specify σ and K.

In order to compute the computational complexity of Ewald summation we ask for the cutoff radius that is an optimal choice, i.e. the choice that solves the Electrostatics problem with error tolerance e^{-p} in minimal time. We consider the number of ions in a cutoff sphere of R:

$$N_c = \frac{4}{3}\pi R^3 \frac{N}{V}$$

where N/V is the number density which is assumed to be uniform. When the time to calculate one interaction pair is t_r, we obtain the execution time

$$T_r \approx \frac{1}{2} N \frac{4}{3}\pi R^3 \frac{N}{V} t_r$$

For the k-space sum we choose wave-vectors according to $\mathbf{k} = \frac{2\pi}{L}(i,j,k)$. The number of index triples in reciprocal space is just:

$$N_k = N_c = \frac{4}{3}\pi K^3 \frac{1}{V_k} = \frac{L^3 K^3}{6\pi^2}$$

where $V_k = (2\pi/L)^3$ is the volume per reciprocal point, and $1/V_k$ the reciprocal point density. We can exploit the symmetry structure of $e^{i\langle \mathbf{r},\mathbf{k}\rangle} = \cos\langle \mathbf{r},\mathbf{k}\rangle + i\sin\langle \mathbf{r},\mathbf{k}\rangle = \cos\langle \mathbf{r},-\mathbf{k}\rangle - i\sin\langle \mathbf{r},-\mathbf{k}\rangle$ in order to execute the sum only over one half-space of $\mathbf{k}$-vectors. Hence, the execution time for the k-space sum is

$$T_k \approx \frac{1}{2}\frac{L^3 K^3}{6\pi^2} t_k$$

In total, we get the time:

$$T_{tot} = \frac{1}{2}\left[N \frac{4}{3}\pi R^3 \frac{N}{V} t_r + \frac{L^3 K^3}{6\pi^2} t_k \right]$$

if we fix the particle density N/V, the accuracy requirement p, and then minimize the execution time with respect to R, we find that:

$$R_{opt} = \left(\frac{p}{\pi}\right)^{1/2} \left(\frac{t_k}{t_r}\right)^{1/6} \frac{L}{N^{1/6}}$$

and the corresponding optimal time is:

$$T_{opt} = 2T_r = 2T_k = \frac{4\pi}{3}\left(\frac{p}{\pi}\right)^{3/2} \sqrt{t_r t_k} N^{3/2}.$$

10.5 Exercise

1. **Conditional convergence**: Compute the interaction energy of the two central charges with the rest of the series:

   $$\cdots + - (+ -) + - \cdots$$

   in a unitless setting (interaction energy $q_i q_j / r$). The charges are $+1$ or -1 and the neighbor distance is 1. Use the two iterative approaches discussed in the script and show their convergence behaviors depending on the number of neighbors considered.

11 Particle Mesh Ewald

The superlinear ($N^{3/2}$) performance of the standard Ewald method comes from the fact that we need to increase both the real-space cutoff and the reciprocal-space cutoff when the system size grows, in order to maintain the same error level. PME seeks to keep the real-space cutoff cutoff constant while maintaining the error level by introducing following trick:

1. Map the charges on a periodic lattice.

2. Solve the Fourier transform on the periodic lattice using FFT.

3. Map the potential back to the charges.

FFT is complete (no cutoff). The FFT-part depends only mildly on N ($N \log N$), while the real-space sum is constant in N when the real-space cutoff is fixed. Thus we arrive at a total performance of $N \log N$.

Mapping the charges to the grid Consider the mesh coordinates

$$\mathbf{r_m} = \frac{m_1}{M_1}\mathbf{a}_1 + \frac{m_2}{M_2}\mathbf{a}_2 + \frac{m_3}{M_3}\mathbf{a}_3$$

with index set

$$\mathbf{m} = (m_1, m_2, m_3)$$

The charge at a mesh point $\mathbf{r_m}$ is in general given by

$$q(\mathbf{r_m}) = \sum_i q_i W(\mathbf{r_m} - \mathbf{r}_i)$$

where W is a spreading function. In PME, we use an approximation to the Gaussian spread functions:

$$q(\mathbf{r_m}) = \sum_{\mathbf{n}} \sum_{j=1}^{N} q_j G_\sigma(\mathbf{r_m} - \mathbf{r}_j + \mathbf{n}L)$$

(Note that for grids of realistic width we must use different spreading functions that preserve the charge $\sum_{\mathbf{m}} q(\mathbf{r_m}) = \sum_i q_i$. We won't go into details about that here)

Solving ϕ^L via the Poisson equation As described in Sec. **??**, we can solve for ϕ either by using a direct solver (via setting up the linear system $\mathbf{Ax} = \mathbf{b}$ and then using a sparse matrix solver), or solve by transforming to reciprocal space, dividing by k^2, and backtransforming to real space via FFT. Either way, when efficiently implemented, the effort is $N \log N$.

We note that the difference from the situation in Sec. **??** is that we do not have point charges but Gaussian spread charges. If these were approximated by a fine grid, we would directly obtain:

$$\hat{\rho}^L_{\mathbf{k_m}} = \frac{1}{V} \mathrm{e}^{-\sigma^2 k^2/2} \sum_{\mathbf{r_m}} q(\mathbf{r_m})\, \mathrm{e}^{-i\langle \mathbf{k_m}, \mathbf{r_m}\rangle}$$

which is transformed into the reciprocal potential:

$$\hat{\phi}^L(\mathbf{k_m}) = \frac{1}{V\epsilon_0} G'(k) \sum_{\mathbf{r_m}} q(\mathbf{r_m})\, \mathrm{e}^{-i\langle \mathbf{k_m}, \mathbf{r_m}\rangle} \tag{21}$$

with

$$G'(k) = \frac{\mathrm{e}^{-\sigma^2 k^2/2}}{k^2} \tag{22}$$

Next, we back-transform $\hat{\phi}$ via inverse FFT:

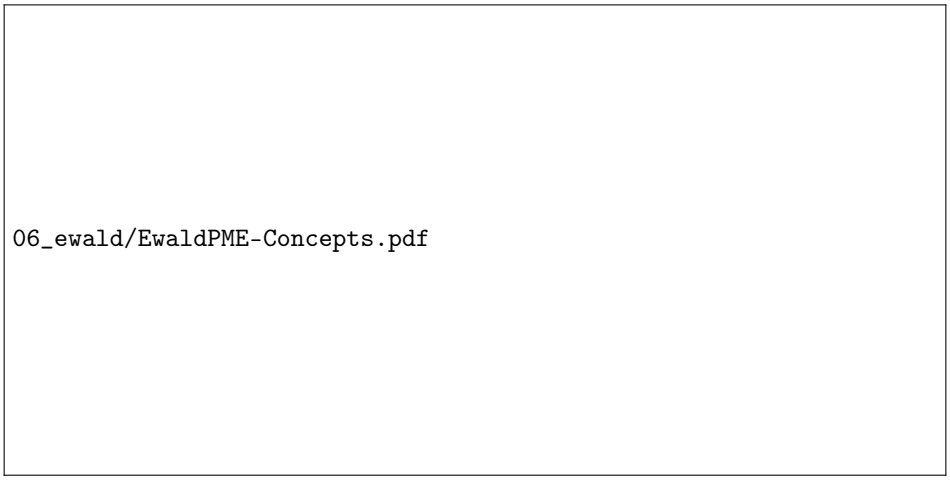
$$\phi^L(\mathbf{r_m}) = \sum_{\mathbf{k_m}} \mathrm{e}^{i\langle \mathbf{k_m}, \mathbf{r_m}\rangle} \hat{\phi}^L(\mathbf{k_m})$$

However, it is more efficient to use a not-so-fine grid and instead of true Gaussians use spread functions with finite support, such that we only need to spread charges to neighboring grid cells. Thus, we introduce discretization errors that can be counteracted by using different coefficients $G'(k)$ instead of Eq. (22). Making these choices appropriately such that computational effort and discretization error are balanced is the art of PME implementations, and we will not go into details about that here.

Back-interpolation We interpolate $\phi^L(\mathbf{r_m})$ back onto the charges, obtaining $\phi^L(\mathbf{r})$ and compute the total energy

$$E^L = \sum_i q_i \phi^L(\mathbf{r}_i)$$

Overview



The computational complexity of PME is $N \log N$ (see Poisson solution, above)

Fast multipole method (FMM)
TODOs:

- Look at **Eric Poisson's Electrodynamics** lecture (chapter 2.10) for multipole expansion of the electric field.

Taken from: VIKAS CHANDRAKANT RAYKAR: A short primer on the fast multipole method

FMM has been voted as one of the 10 most important algorithms.

FMM allows the matrix-vector products with special matrix structures to be computed in $\mathcal{O}(N)$ or $\mathcal{O}(N \log N)$ operations, while the direct implementation costs $\mathcal{O}(N^2)$. Originally, the method was proposed fro gravitational or electrostatic potential problems in two or three dimensions, leading to the name of the algorithm, but the algorithm is more general.

12 Potentials and factorization

We are interesting in computing the following sum:

$$v_j = \sum_{i=1}^{N} q_i \Phi(y_j, x_i) \quad j = 1, ..., M \tag{23}$$

where $\{x_i \in \mathbb{R}^d\}_{i=1,...,N}$ are called source points, $\{y_i \in \mathbb{R}^d\}_{i=1,...,M}$ are called target points, $\{q_i \in \mathbb{R}^d\}_{i=1,...,N}$ are called source weights, and Φ is the potential function. $\Phi(y_j, x_i)$ is the contribution of source at x_i towards the target point y_j. The computational complexity to evaluate (23) directly is $\mathcal{O}(NM)$. The FMM provides the approximate solution to (23) within a user-defined error level in a lower computational complexity. Compared to the FFT, the FMM does not requie the data to be uniformly sampled.

TODO: Does the FFT also solve the above problem? Need to work this out.

We refer to $\Phi(y, x_i)$ as the *field* or *potential* of the ith unit source. We mainly focus on scalar fields. $\Phi(y, x_i)$. Examples:

Gravity (singular at $x = x_i$):

$$\Phi(y, x_i) = \frac{1}{\|y - x_i\|}$$

Gaussian field (regular everywhere):

$$\Phi(y, x_i) = \mathrm{e}^{-\|y-x_i\|^2/h^2}$$

12.1 Near-field R-expansion

For any $x_* \in \mathbb{R}^d$ we call the expansion

$$\Phi(y, x_i) = \sum_{m=0}^{\infty} a_m(x_i, x_*)\, R_m(y - x_*) \tag{24}$$

regular (local) inside a sphere $B^<_{r_*}(x_*) = \{y \in \mathbb{R}^d : \|y - x_*\| < r_*\}$, if the series converges for all $y \in B^<_{r_*}(x_*)$. $a_m(x_i, x_*)$ are the *expansion coefficients* and $R_m(y - x_*)$ are the *regular basis functions*. x_* is called the *expansion center*. Eq. (24) is called the *R-expansion* of the potential $\Phi(y, x_i)$.

For a regular potential, the convergence region $B^<_{r_*}(x_*)$ can be arbitrary (the R-expansion is valid for all three cases shown in Fig. 2), but for a singular potential, $B^<_{r_*}(x_*)$ cannot include the singular point (only Fig. 2a, b are valid) (**Note: that would mean that the expansion cannot be made around the sources, e.g. charges...**).

Example 1: R-expansion of the one-dimensional Gaussian

Consider the one-dimensional regular function

$$\Phi(y, x_i) = \mathrm{e}^{-(y-x_i)^2/h^2}$$

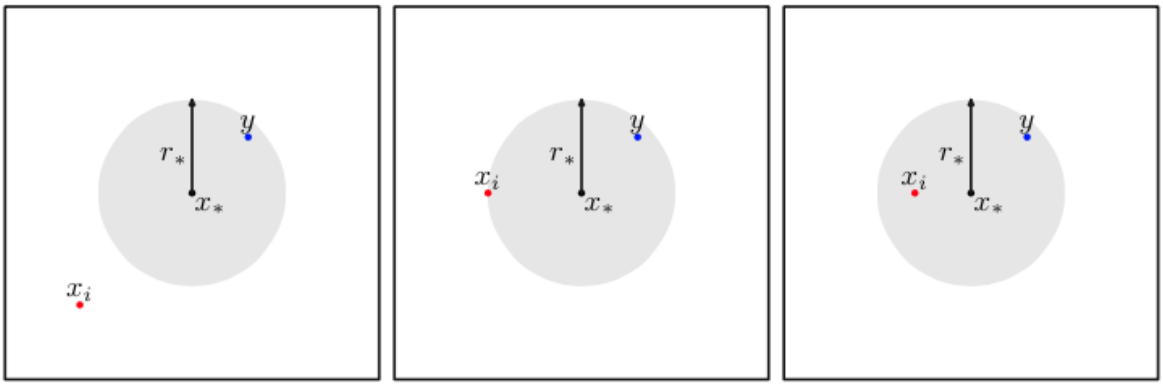


Figure 1: Near-field (local R-) expansion for (a) $r_* < \|x_i - x_*\|$, (b) $r_* = \|x_i - x_*\|$, (c) $r_* > \|x_i - x_*\|$

For any $x_* \in \mathbb{R}$ and $\|y - x_*\| < r_* < \infty$. Using the Taylor expansion, the R-expansion can be written as:

$$\begin{aligned}
\Phi(y,\, x_i) &= \mathrm{e}^{-(y+x_*-x_*-x_i)^2/h^2} \\
&= \mathrm{e}^{-((y-x_*)-(x_i-x_*))^2/h^2} \\
&= \mathrm{e}^{-(y-x_*)^2/h^2}\mathrm{e}^{-(x_i-x_*)^2/h^2}\mathrm{e}^{2(y-x_*)(x_i-x_*)/h^2}
\end{aligned}$$

we can then use the series expression for the mixed exponential:

$$\mathrm{e}^{2(y-x_*)(x_i-x_*)/h^2} = \sum_{m=0}^{\infty} \frac{2^m}{m!} \left(\frac{y-x_*}{h}\right)^m \left(\frac{x_i-x_*}{h}\right)^m$$

and rewrite the potential towards the expansion:

$$\begin{aligned}
\Phi(y,\, x_i) &= \sum_{m=0}^{\infty} \frac{2^m}{m!} \mathrm{e}^{-(x_i-x_*)^2/h^2} \left(\frac{x_i-x_*}{h}\right)^m \mathrm{e}^{-(y-x_*)^2/h^2} \left(\frac{y-x_*}{h}\right)^m \\
&= \sum_{m=0}^{\infty} a_m(x_i,\, x_*)\, R_m(y-x_*)
\end{aligned}$$

with coefficients and basis functions given by:

$$\begin{aligned}
a_m(x_i,\, x_*) &= \frac{2^m}{m!} \mathrm{e}^{-(x_i-x_*)^2/h^2} \left(\frac{x_i-x_*}{h}\right)^m \\
R_m(y-x_*) &= \mathrm{e}^{-(y-x_*)^2/h^2} \left(\frac{y-x_*}{h}\right)^m
\end{aligned}$$

Example 2: R-expansion of a one-dimensional graviational field
The one-dimensional gravitational field which is singular at $y = x_i$:

$$\Phi(y,\, x_i) = \frac{1}{y - x_i}$$

For any $x_* \in \mathbb{R}$ and $\|y - x_*\| < r_* \leq \|x_i - x_*\|$ (to make sure we do not include the singular point) **Note: understand this better**.

$$\begin{aligned}
\Phi(y,\, x_i) &= \frac{1}{y - x_* + x_* - x_i} \\
&= -\frac{1}{(x_i - x_*)\left(1 - \frac{y-x_*}{x_i-x_*}\right)} \\
&= -\frac{1}{(x_i - x_*)} \left(1 - \frac{y-x_*}{x_i-x_*}\right)^{-1}
\end{aligned}$$

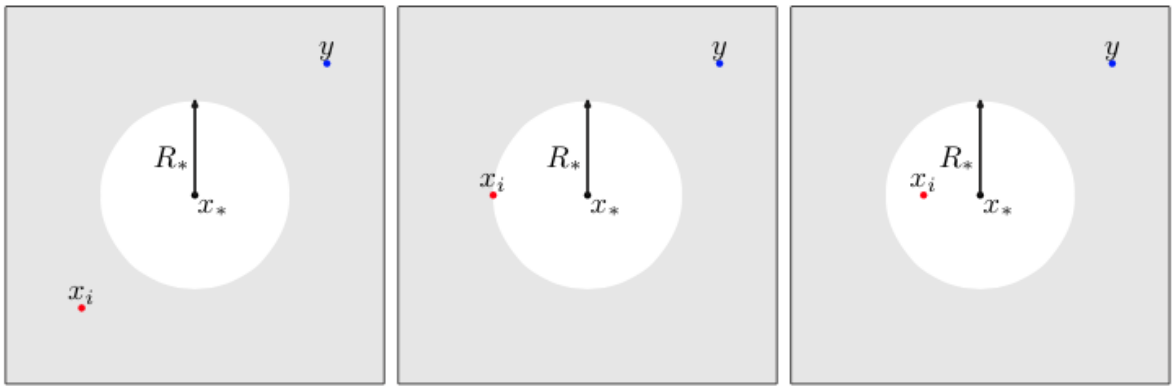


Figure 2: Far-field S-expansion (a) $R_* < \|x_i - x_*\|$, (b) $R_* = \|x_i - x_*\|$, (c) $R_* > \|x_i - x_*\|$

We use the geometric progression. Since $\|y - x_*\| < \|x_i - x_*\|$:

$$\left(1 - \frac{y - x_*}{x_i - x_*}\right)^{-1} = \sum_{m=0}^{\infty} \frac{(y - x_*)^m}{(x_i - x_*)^m}$$

obtaining the expansion

$$\Phi(y,\, x_i) = \sum_{m=0}^{\infty} a_m(x_i,\, x_*)\, R_m(y - x_*)$$

with

$$\begin{aligned} a_m(x_i,\, x_*) &= -\frac{1}{(x_i - x_*)^{m+1}} \\ R_m(y - x_*) &= (y - x_*)^m \end{aligned}$$

12.2 Far-field S-expansion

For any $x_* \in \mathbb{R}^d$ we call the expansion

$$\Phi(y,\, x_i) = \sum_{m=0}^{\infty} b_m(x_i,\, x_*)\, S_m(y - x_*) \tag{25}$$

far-field expansion outside a sphere sphere $B^{>}_{R_*}(x_*) = \{y \in \mathbb{R}^d : \|y - x_*\| > R_*\}$, if the series converges for all $y \in B^{>}_{R_*}(x_*)$. $b_m(x_i,\, x_*)$ are the *expansion coefficients* and $S_m(y - x_*)$ are basis functions (which can be singular at $y = x_*$). Eq. (25) is called the *S-expansion* of the potential $\Phi(y,\, x_i)$. For a regular potential, the convergence region $B^{>}_{R_*}(x_*)$ can be arbitrary (the S-expansion is valid for all three cases shown in Fig. 2), but for a singular potential, $B^{>}_{R_*}(x_*)$ cannot include the singular point (only Fig. 2b, c are valid).

Example 1: S-expansion of the one-dimensional Gaussian

For a regular function, the R-expansion can be used as S-expansion since there are no singular points. However, in order to get a rapidly converging series a different choice is better. Consider again

$$\Phi(y,\, x_i) = \mathrm{e}^{-(y - x_i)^2/h^2}$$

For any $x_* \in \mathbb{R}$ and $\|y - x_*\| > R_* < \infty$:

$$\Phi(y,\, x_i) = \sum_{m=0}^{\infty} b_m(x_i,\, x_*)\, S_m(y - x_*)$$

with coefficients and basis functions given by:

$$\begin{aligned} b_m(x_i,\, x_*) &= \frac{2^m}{m!}\mathrm{e}^{-(x_i-x_*)^2/h^2}\,(x_i-x_*)^m \\ S_m(y-x_*) &= \mathrm{e}^{-(y-x_*)^2/h^2}\,(y-x_*)^m \end{aligned}$$

The only difference from the R-expansion is that we don't have the factor h^{-1} inside the last factor. **Note: understand this difference**.

Example 2: S-expansion of a one-dimensional graviational field

Consider again the one-dimensional gravitational field which is singular at $y = x_i$:

$$\Phi(y,\, x_i) = \frac{1}{y-x_i}$$

For any $x_* \in \mathbb{R}$ and $\|y - x_*\| > R_* \geq \|x_i - x_*\|$ (to make sure we do not include the singular point x_i):

$$\begin{aligned} \Phi(y,\, x_i) &= \frac{1}{y-x_*+x_*-x_i} \\ &= \frac{1}{(y-x_*)\left(1-\frac{x_i-x_*}{y-x_*}\right)} \\ &= \frac{1}{(y-x_*)}\sum_{m=0}^{\infty}\frac{(x_i-x_*)^m}{(y-x_*)^m} \\ &= \sum_{m=0}^{\infty}\frac{(x_i-x_*)^m}{(y-x_*)^{m+1}} \end{aligned}$$

arriving at

$$\Phi(y,\, x_i) = \sum_{m=0}^{\infty} b_m(x_i,\, x_*)\, S_m(y-x_*)$$

with coefficients and basis functions given by:

$$\begin{aligned} b_m(x_i,\, x_*) &= (x_i-x_*)^m \\ S_m(y-x_*) &= \frac{1}{(y-x_*)^{m+1}} \end{aligned}$$

Why R and S-expansions? If the potential has a singular point x_i, then we use the R-expansion for all $\|y - x_*\| < \|x_i - x_*\|$ and the S-expansion for all $\|y - x_*\| > \|x_i - x_*\|$. The singular point is at the boundary of the regions for the R- and S-expansion. Also in the case of regular pontentials, depending upon how far y is from x_* either the S-expansion or the R-expansion will converge much more rapidly than the other.

12.3 Middleman method

Regular potentials

Consider a potential that is regular everywhere. We use the R-expansion about x_* for factorizing the potential. The field at y_j can then be evaluated as:

$$\begin{aligned} v_j &= \sum_{i=1}^{N} q_i \Phi(y_j,\, x_i) \\ &= \sum_{i=1}^{N} q_i \left[\sum_{m=0}^{\infty} a_m(x_i,\, x_*)\, R_m(y_j - x_*)\right] \\ &\approx \sum_{i=1}^{N} q_i \left[\sum_{m=0}^{p-1} a_m(x_i,\, x_*)\, R_m(y_j - x_*)\right]. \end{aligned}$$

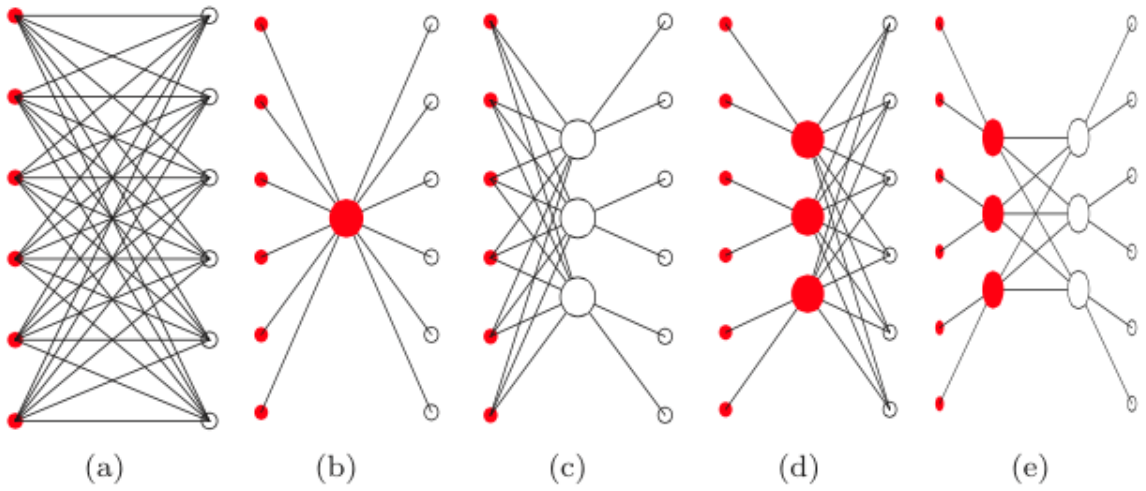


Figure 3: (a) Direct computation (b) Middleman method (c) middleman with target clusters (d) middleman with source clusters (e) Single level FMM.

The truncation number p is selected based on the desired error. One of the trickiest part in designing a good FMM algorithm lies in getting a good error bound for the residual error. The approximated potential is given by:

$$\begin{aligned} v_j &= \sum_{i=1}^{N} q_i \left[\sum_{m=0}^{p-1} a_m(x_i,\, x_*)\, R_m(y_j - x_*) \right] \\ &= \sum_{m=0}^{p-1} \left[\sum_{i=1}^{N} q_i a_m(x_i,\, x_*) \right] R_m(y_j - x_*) \\ &= \sum_{m=0}^{p-1} A_m R_m(y_j - x_*) \end{aligned}$$

where $A_m = \sum_{i=1}^{N} q_i a_m(x_i,\, x_*)$, which depends only on the source, can be computed in one pass for different m. The computational complexity for computing all relevant A_m is $\mathcal{O}(pN)$, and the complexity for computing all $\hat{v}_j$ is $\mathcal{O}(pM)$, in total we have $\mathcal{O}(pN + pM)$. As long as $p \ll \min(M,\, N)$ we have a reduction in complexity.

This method is called middleman method because we expand the potential around a single expansion center x_* (Fig. 3b). However, we can use more than one expansion center (Fig. 3d).

Singular potentials

For singular potentials, we can use the middleman method, if the source and the target points are well separated (Fig. 4). If the sources and targets are not well separated, we do a space partitioning. For example, we partition with respect to the target points. For each target cluster we use the R-expansion for all sources outside a neighborhood of the cluster.

There will be some sources which happen to lie within the neighborhood of these clusters, where the potential is singular at each of these source points. If the number of such sources is few, we can directly sum their contributions (direct sum for nearest neighbors, similar to Ewald method). If there are many, this leads to the idea of a single level FMM, discussed next.

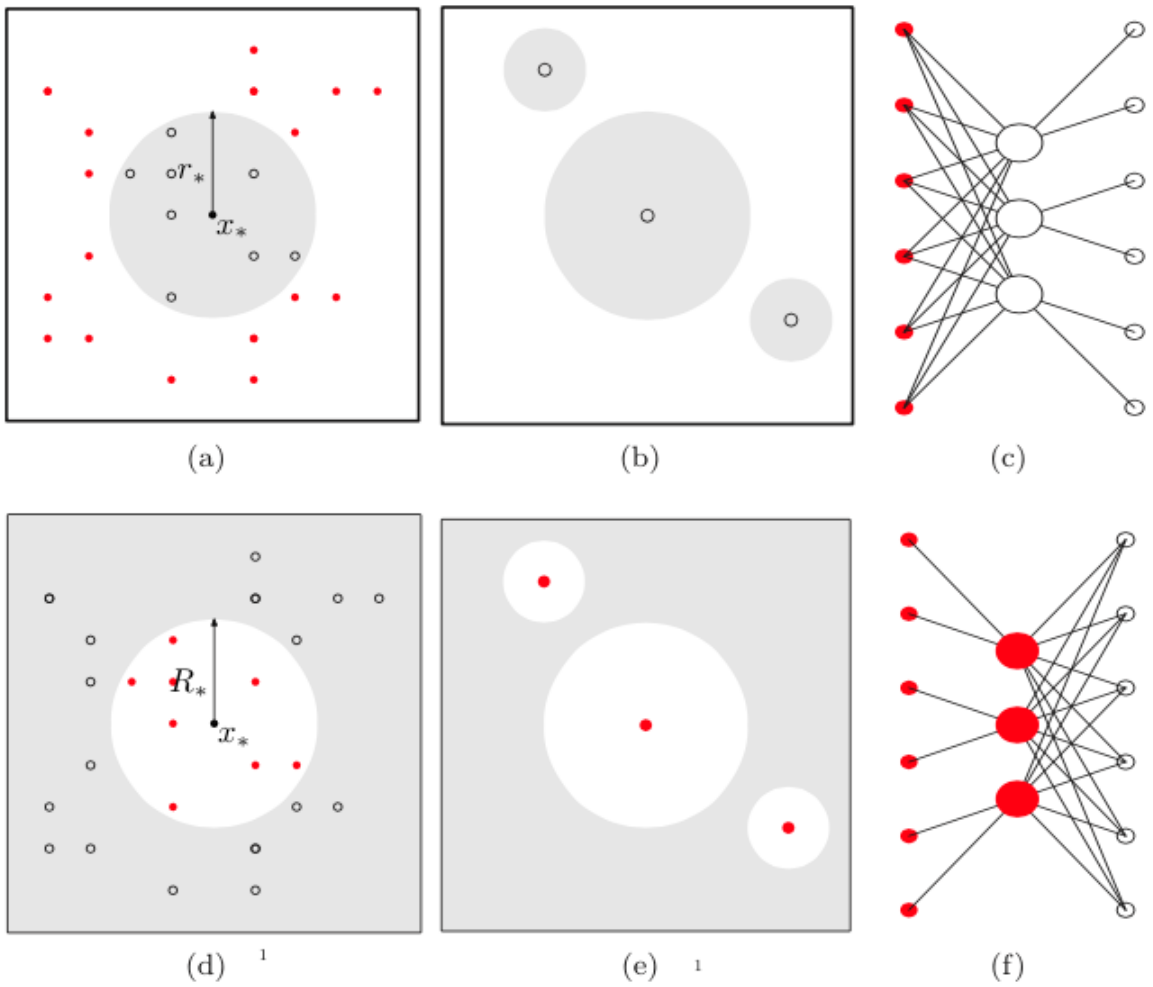


Figure 4: For singular potentials we can use the middleman approach if the source and the target points are well separated. In (a) and (b) we will use the R-expansion at the center of each target clusters and in (d) and (e) we will use the S-expansion at the center of each source clusters. The red filled circles are the source points and the back circles are the target points.

12.4 Translations

Let $\{F_n(y - x_{*1})\}_{n=0}^{\infty}$ and $\{G_m(y - x_{*2})\}_{m=0}^{\infty}$ be two sets of basis functions centered at x_{*1} and x_{*2} such that $\Phi(y,\, x_i)$ can be represented by two uniformly and absolutely convergent time series as:

$$\begin{aligned}\Phi(y,\, x_i) &= \sum_{n=0}^{\infty} a_n(x_i,\, x_{*1}) F_n(y - x_{*1}) \quad \forall y \in \Omega_1 \subset \mathbb{R}^d \\ \Phi(y,\, x_i) &= \sum_{m=0}^{\infty} b_m(x_i,\, x_{*2}) G_m(y - x_{*2}) \quad \forall y \in \Omega_2 \subset \Omega_1\end{aligned}$$

Thus, we have two different representations for points in a smaller and a larger neighborhood. We call $(F \mid G)(t)$ a *translation operator* which relates two sets of coefficients:

$$\{b_m(x_i,\, x_{*2})\} = (F \mid G)(t)\{a_n(x_i,\, x_{*1})\}, \quad t = x_{*2} - x_{*1}.$$

$R \mid R$ re-expansion (local to local)

Consider the R-expansion of a potential $\Phi(y, x_i)$ about the expansion center $x_* \in \mathbb{R}^d$:

$$\Phi(y,\, x_i) = \sum_{m=0}^{\infty} a_m(x_i,\, x_*) R_m(y - x_*) \quad \forall y \in B_r^<(x_*)$$

We are interested in writing the same function of a shifted basis $R_n(y - (x_* + t))$ about the shifted expansion center $x_* + t \in B_r^<(x_*)$ that is still local.

$$\Phi(y,\, x_i) = \sum_{n=0}^{\infty} \tilde{a}_n(x_i,\, x_* + t) R_n(y - (x_* + t)) \quad \forall y \in B_{r1}^<(x_* + t) \subset B_r^<(x_*)$$

where $B_{r1}^<(x_* + t)$ is a sub-neighborhood within $B_r^<(x_*)$. The matrix operator $(R \mid R)(t)$ relates the coefficiens $a_m(x_i,\, x_*)$ and $\tilde{a}_n(x_i,\, x_* + t)$ as,

$$\tilde{a}_n(x_i,\, x_* + t) = \sum_{m=0}^{\infty} (R \mid R)(t) a_m(x_i,\, x_*)$$

$S \mid S$ re-expansion (multipole to multipole)

$S \mid R$ re-expansion (multipole to local)

$R \mid S$ re-expansion (local to multipole)

12.5 Single level FMM

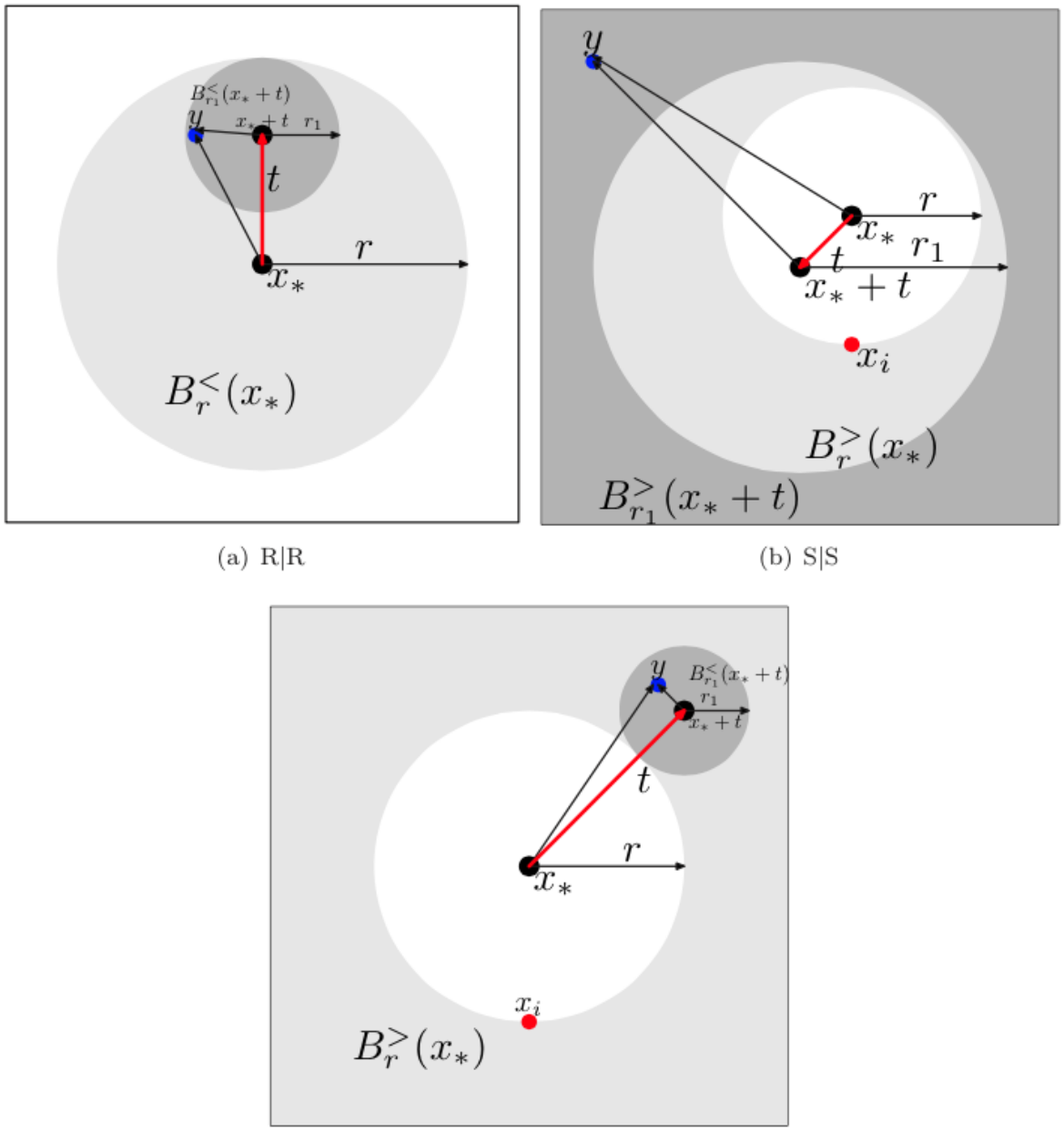
Problem with the middleman method are:

- For a regular potential the expansion is valid everywhere, but the truncation number can be huge for large domains to provide the desired accuracy, resulting in no saving.
- For singular potentials the middleman approach can be used only when either the source or targets or naturally grouped

Space subdivision

Given the above series expressions and translation operators, the single level FMM proceeds as follows. First the space is partitioned into boxes. Each box will contain a certain number of sources and targets. For any box n we define the following three spatial domains:

$$\begin{aligned}I_1(n) &= n \\ I_2(n) &= \{\text{Neighbors}(n)\} \cup \{n\} \\ I_3(n) &= \{\text{All boxes}\} - \{I_2(n)\}\end{aligned}$$



(a) R|R

(b) S|S

(c) S|R

Figure 5: Schematic illustrating (a) R|R re-expansion, (b) S|S re-expansion, and (c) S|R re-expansion.

Let the target points be $y \in n$. We define the following three potentials

$$\begin{aligned}
\Phi_1^n(y) &= \sum_{x_i \in I_1(n)} q_i \Phi(y,\, x_i) \\
\Phi_2^n(y) &= \sum_{x_i \in I_2(n)} q_i \Phi(y,\, x_i) \\
\Phi_3^n(y) &= \sum_{x_i \in I_3(n)} q_i \Phi(y,\, x_i)
\end{aligned}$$

Due to the definition of the domains we can write the total potential as:

$$\Phi(y) = \Phi_2^n(y) + \Phi_3^n(y)$$

R expansion

The potential of the outside sources onto the central point $y \in n$, $\Phi_3^n(y)$, has to be computed using the local R expansion of $\Phi(y,\, x_i)$ about the box center x_c^n which is near y. If y is close to x_c^n then the local R-expansion series will converge quickly.

Exchanging the order of summation and consolidating source terms we have:

$$\begin{aligned}
\Phi_3^n(y) &= \sum_{x_i \in I_3(n)} q_i \Phi(y,\, x_i) \\
&= \sum_{x_i \in I_3(n)} q_i \left[\sum_{m=0}^{\infty} a_m(x_i,\, x_c^n) R_m(y - x_c^n) \right] \\
&= \sum_{m=0}^{\infty} \left[\sum_{x_i \in I_3(n)} q_i a_m(x_i,\, x_c^n) \right] R_m(y - x_c^n) \\
&= \sum_{m=0}^{\infty} A_m^n R_m(y - x_c^n)
\end{aligned}$$

with $A_m^n = \sum_{x_i \in I_3(n)} q_i a_m(x_i,\, x_c^n)$.

S expansion

If y is far away from x_c^n, then the R-expansion may converge very slowly or not at all. We need a larger radius of convergence so that the number of boxes are small. For this reason we use the S-expansions and translate it to an R-expansion via the $S \mid R$ translation operator. The S-expansions are computed about the box centers for all the boxes. Exchanging the order of summation and cosolidating the source terms we have $\Phi_1^l(y)$ the potential at $y \in I_3(l)$ due to all sources $x_i \in I_1(l)$:

$$\begin{aligned}
\Phi_1^l(y) &= \sum_{x_i \in I_1(l)} q_i \Phi(y,\, x_i) \\
&= \sum_{x_i \in I_1(l)} q_i \left[\sum_{m=0}^{\infty} b_m(x_i,\, x_c^l) S_m(y - x_c^l) \right] \\
&= \sum_{m=0}^{\infty} \left[\sum_{x_i \in I_1(l)} q_i b_m(x_i,\, x_c^l) \right] S_m(y - x_c^l) \\
&= \sum_{m=0}^{\infty} B_m^l S_m(y - x_c^l)
\end{aligned}$$

with $B_m^l = \sum_{x_i \in I_1(l)} q_i b_m(x_i,\, x_c^l)$

$S \mid R$ translation

Now we want to write the potential $\Phi_1^l(y)$ (expanded using an S expansion around x_c^l as an R-expansion around x_c^n.

$$\begin{aligned}
\Phi_1^l(y) &= \sum_{x_i \in I_1(l)} q_i \left[\sum_{m=0}^{\infty} b_m(x_i,\, x_c^l) S_m(y - x_c^l) \right] \\
&= \sum_{x_i \in I_1(l)} q_i \left[\sum_{k=0}^{\infty} \tilde{a}_k(x_i,\, x_c^l + (x_c^n - x_c^l))\, R_k(y - (x_c^l + (x_c^n - x_c^l))) \right] \\
&= \sum_{k=0}^{\infty} \left[\sum_{x_i \in I_1(l)} q_i \tilde{a}_k(x_i,\, x_c^l + (x_c^n - x_c^l)) \right] R_k(y - x_c^n) \\
&= \sum_{k=0}^{\infty} \left[\sum_{m=0}^{\infty} (S \mid R)_{km}(x_c^n - x_c^l) \left\{ \sum_{x_i \in I_1(l)} q_i b_m(x_i,\, x_c^l) \right\} \right] R_k(y - x_c^n) \\
&= \sum_{k=0}^{\infty} \left[\sum_{m=0}^{\infty} (S \mid R)_{km}(x_c^n - x_c^l) B_m^l \right] R_k(y - x_c^n) \\
&= \sum_{k=0}^{\infty} \tilde{A}_k^{nl} R_k(y - x_c^n)
\end{aligned}$$

where $\tilde{A}_k^{nl} = \sum_{m=0}^{\infty} (S \mid R)_{km}(x_c^n - x_c^l) B_m^l$. The R-expansions about the box center x_c^n are given by $\Phi_3^n(y)$ where we can write $A_m^n = \sum_{x_i \in I_3(n)} q_i a_m(x_i, x_c^n) = \sum_{l \in I_e(n)} \tilde{A}_m^{nl}$. So we have R-expansions about the box centers and very few points (in $I_2(n)$) for which valid expansions could not be constructed. These are evaluated directly and added to the R-expansions evaluated at the evaluation points. The speedup is achieved by appropriately truncating each series.

13 Multipole expansion for the electrostatic field

14 Multipole expansion for the gravity field

From: https://en.wikipedia.org/wiki/Gravitational_potential

The gravitational potential at a point $\mathbf{x}$ is given by

$$V(\mathbf{x}) = -\int_{\mathbb{R}^3} \frac{G}{|\mathbf{x} - \mathbf{r}|}\, \mathrm{d}m(\mathbf{r}).$$

The potential can be expanded in a series of Legendre polynomials. Represent the points $\mathbf{x}$ and $\mathbf{r}$ as position vectors relative to the center of mass. The denominator in the integral is expressed as the square root of the square to give

$$\begin{aligned}
V(\mathbf{x}) &= -\int_{\mathbb{R}^3} \frac{G}{\sqrt{|\mathbf{x}|^2 - 2\mathbf{x}\cdot\mathbf{r} + |\mathbf{r}|^2}}\, \mathrm{d}m(\mathbf{r}) \\
&= -\frac{1}{|\mathbf{x}|} \int_{\mathbb{R}^3} G \Bigg/ \sqrt{1 - 2\frac{r}{|\mathbf{x}|}\cos\theta + \left(\frac{r}{|\mathbf{x}|}\right)^2}\, \mathrm{d}m(\mathbf{r})
\end{aligned}$$

where in the last integral, $r = |\mathbf{r}|$ and θ is the angle between $\mathbf{x}$ and $\mathbf{r}$.

The integrand can be expanded as a Taylor series in $\mathbf{Z} = \mathbf{r}/|\mathbf{x}|$, by explicit calculation of the coefficients. A less laborious way of achieving the same result is by using the generalized binomial theorem. The resulting series is the generating function for the Legendre polynomials:

$$\left(1 - 2XZ + Z^2\right)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} Z^n P_n(X)$$

Algorithm 1: Single Level Fast Multipole Method

Input :

$x_i \in \mathbf{R}^d \ i = 1, \ldots, N$ `/* N sources in d dimensions. */`
$q_i \in \mathbf{R}^+ \ i = 1, \ldots, N$ `/* source weights. */`
$y_j \in \mathbf{R}^d \ j = 1, \ldots, M$ `/* M targets in d dimensions. */`
$\Phi(y, x_i)$ `/* potential at` y `due to source` x_i`. */`
$\epsilon > 0$ `/* Desired error. */`

Output: Computes an approximation $\hat{\Phi}(y_j)$ to $\Phi(y_j) = \sum_{i=1}^{N} q_i \Phi(y_j, x_i)$ such that the $|\hat{\Phi}(y_j) - \Phi(y_j)| \leq \epsilon$.

Step 0 *Partition the space into boxes.*;

Step 1 *Decide the truncation numbers* p_1 *and* p_2 *such that the error is* $\leq \epsilon$;

Step 2 *S- and R-expansion of the potential function about each box center.*
$\forall x_i \in I_1(l) \ \forall y \in I_3(l) \ \Phi(y, x_i) = \sum_{r=0}^{p_1 - 1} b_r(x_i, x_c^l) S_r(y - x_c^l)$
$\forall x_i \in I_3(l) \ \forall y \in I_1(l) \ \Phi(y, x_i) = \sum_{q=0}^{p_2 - 1} a_q(x_i, x_c^l) R_q(y - x_c^l)$;

Step 3 *Consolidate the S-expansion coefficients for each box.*
$B_r^l = \sum_{x_i \in I_1(l)} q_i b_r(x_i, x_c^l)$ *for* $r = 0, \ldots, p_1 - 1$;

Step 4 *S|R translate the expansion coefficients.*
$\widetilde{A_q^{nl}} = \sum_{r=0}^{p_1 - 1} (S|R)_{qr}(x_c^n - x_c^l) B_r^l$ *for* $q = 0, \ldots, p_2 - 1$
$A_q^n = \sum_{l \in I_3(n)} \widetilde{A_q^{nl}}$ *for* $q = 0, \ldots, p_2 - 1$;

Step 5 *Final summation.*
$\Phi(y \in I_1(n)) = \sum_{x_i \in I_2(n)} q_i \Phi(y, x_i) + \sum_{q=0}^{p_2 - 1} A_q^n R_q(y - x_c^n)$;

Figure 6: Single-level FMM

valid for $|X| \leq 1$ and $|Z| < 1$. The coefficients P_n are the Legendre polynomials of degree n. Therefore, the Taylor coefficients of the integrand are given by the Legendre polynomials in $X = \cos\theta$. So the potential can be expanded in a series that is convergent for positions $\mathbf{x}$ such that $r < |\mathbf{x}|$ for all mass elements of the system (i.e., outside a sphere, centered at the center of mass, that encloses the system):

$$\begin{aligned} V(\mathbf{x}) &= -\frac{G}{|\mathbf{x}|} \int \sum_{n=0}^{\infty} \left(\frac{r}{|\mathbf{x}|}\right)^n P_n(\cos\theta)\, \mathrm{d}m(\mathbf{r}) \\ &= -\frac{G}{|\mathbf{x}|} \int \left(1 + \left(\frac{r}{|\mathbf{x}|}\right)\cos\theta + \left(\frac{r}{|\mathbf{x}|}\right)^2 \frac{3\cos^2\theta - 1}{2} + \cdots\right) \mathrm{d}m(\mathbf{r}) \end{aligned}$$

The integral $\int r\cos\theta\, \mathrm{d}m$ is the component of the center of mass in the $\mathbf{x}$ direction; this vanishes because the vector x emanates from the center of mass. So, bringing the integral under the sign of the summation gives

$$V(\mathbf{x}) = -\frac{GM}{|\mathbf{x}|} - \frac{G}{|\mathbf{x}|} \int \left(\frac{r}{|\mathbf{x}|}\right)^2 \frac{3\cos^2\theta - 1}{2}\, \mathrm{d}m(\mathbf{r}) + \cdots$$

This shows that elongation of the body causes a lower potential in the direction of elongation, and a higher potential in perpendicular directions, compared to the potential due to a spherical mass, if we compare cases with the same distance to the center of mass. (If we compare cases with the same distance to the surface the opposite is true.)