Topic: Modeling and Numerical Simulation for Rarefied Gases

Lecture 2: Numerical Methods for the Boltzmann equation

Zhenning Cai

Department of Mathematics, NUS



Summer School at Tianjin University, China

1 Jul 2026

1 Introduction to the Fourier spectral method

2 Fourier spectral method for the Boltzmann collision operator

3 Fix the physical properties

4 Generalization to spatially inhomogeneous Boltzmann equation

5 Summary

Function approximation in the frequency space

Let $f(\boldsymbol{v})$ be a function defined on $\mathbb{R}^3$. Since

$$f(\boldsymbol{v}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \hat{f}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}\cdot\boldsymbol{v}} d\boldsymbol{\xi},$$

Function approximation in the frequency space

Let $f(\boldsymbol{v})$ be a function defined on $\mathbb{R}^3$. Since

$$f(\boldsymbol{v}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \hat{f}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}\cdot\boldsymbol{v}} d\boldsymbol{\xi},$$

we can approximate $f(\boldsymbol{v})$ by

$$f(\boldsymbol{v}) \approx f_N(\boldsymbol{v}) := \frac{1}{(2\pi)^3} \sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} \sum_{k_3=-N}^{N} \hat{f}(\Delta\xi \boldsymbol{k}) \mathrm{e}^{\mathrm{i}\Delta\xi \boldsymbol{k}\cdot\boldsymbol{v}} \Delta\xi^3,$$

Function approximation in the frequency space

Let $f(\boldsymbol{v})$ be a function defined on $\mathbb{R}^3$. Since

$$f(\boldsymbol{v}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \hat{f}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}\cdot\boldsymbol{v}} d\boldsymbol{\xi},$$

we can approximate $f(\boldsymbol{v})$ by

$$f(\boldsymbol{v}) \approx f_N(\boldsymbol{v}) := \frac{1}{(2\pi)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}(\boldsymbol{\xi}_{\boldsymbol{k}}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}_{\boldsymbol{k}}\cdot\boldsymbol{v}} \Delta\xi^3,$$

where $\boldsymbol{\xi}_{\boldsymbol{k}} = \Delta\xi \boldsymbol{k}$.

Function approximation in the frequency space

Let $f(\boldsymbol{v})$ be a function defined on $\mathbb{R}^3$. Since

$$f(\boldsymbol{v}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \hat{f}(\boldsymbol{\xi}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}\cdot\boldsymbol{v}} d\boldsymbol{\xi},$$

we can approximate $f(\boldsymbol{v})$ by

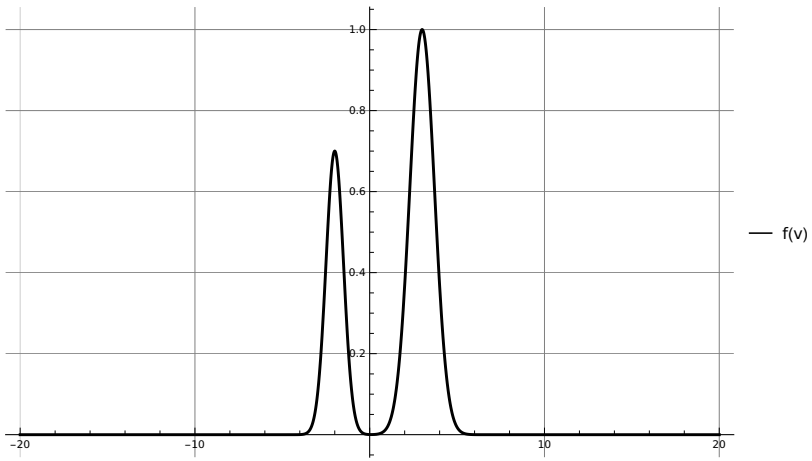
$$f(\boldsymbol{v}) \approx f_N(\boldsymbol{v}) := \frac{1}{(2\pi)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}(\boldsymbol{\xi}_{\boldsymbol{k}}) \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}_{\boldsymbol{k}}\cdot\boldsymbol{v}} \Delta\xi^3,$$

where $\boldsymbol{\xi}_{\boldsymbol{k}} = \Delta\xi \boldsymbol{k}$.

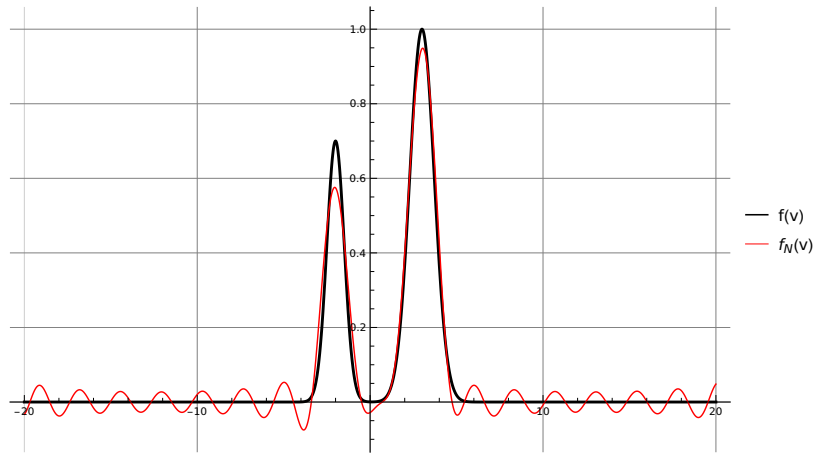
However, f_N is a periodic function, while f is not.

Example

- $f(v) = \exp(-(v-3)^2) + 0.7\exp(-2(v+2)^2)$
- $\hat{f}(\xi) = \sqrt{\pi}\,\mathrm{e}^{-\xi^2/4}\,\mathrm{e}^{-3\mathrm{i}\xi} + 0.7\sqrt{\pi/2}\,\mathrm{e}^{-\xi^2/8}\,\mathrm{e}^{2\mathrm{i}\xi}$



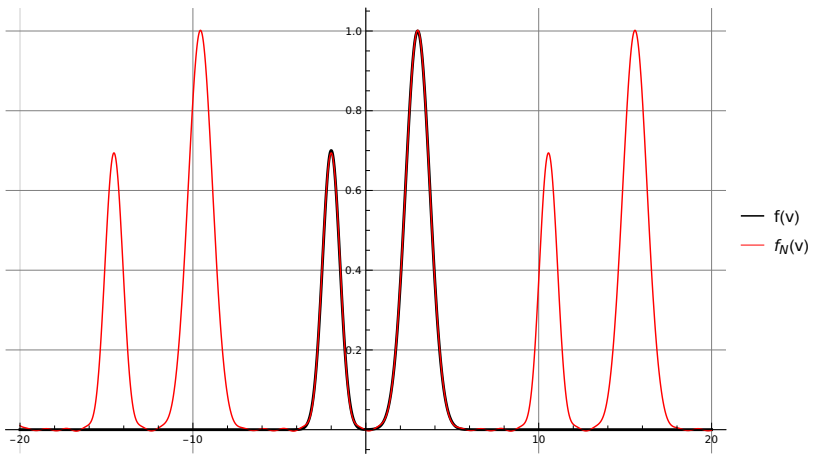
(a) $f(v)$



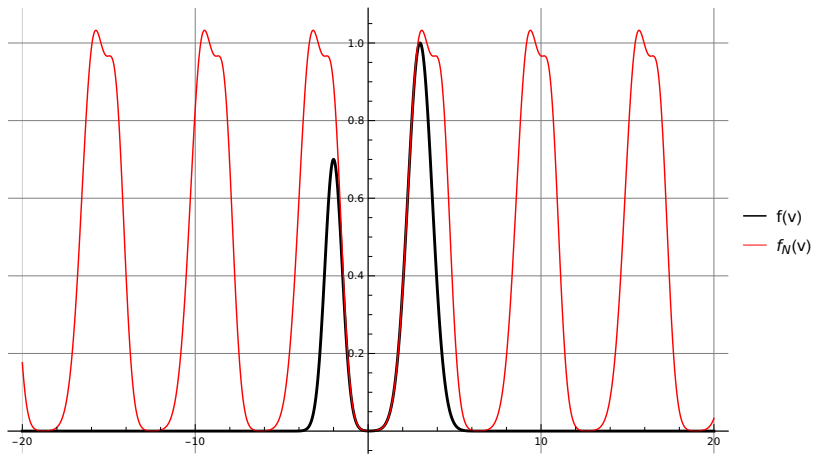
(b) $\Delta\xi = 0.25$

Example

- $f(v) = \exp(-(v-3)^2) + 0.7\exp(-2(v+2)^2)$
- $\hat{f}(\xi) = \sqrt{\pi}\,\mathrm{e}^{-\xi^2/4}\,\mathrm{e}^{-3\mathrm{i}\xi} + 0.7\sqrt{\pi/2}\,\mathrm{e}^{-\xi^2/8}\,\mathrm{e}^{2\mathrm{i}\xi}$



(c) $\Delta\xi = 0.5$



(d) $\Delta\xi = 1.0$

Discrete Fourier transform (DFT)

For

$$f_N(\boldsymbol{v}) = \frac{1}{(2\pi)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\boldsymbol{\xi}_{\boldsymbol{k}}\cdot\boldsymbol{v}} \Delta\xi^3,$$

the period is $[-L, L]^3$ with $L = \pi/\Delta\xi$.

Discrete Fourier transform (DFT)

For

$$f_N(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{v}/L},$$

the period is $[-L, L]^3$.

Discrete Fourier transform (DFT)

For

$$f_N(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{v}/L},$$

the period is $[-L, L]^3$. The coefficients $\hat{f}_{\boldsymbol{k}}$ can be computed by

$$\hat{f}_{\boldsymbol{k}} = \frac{1}{c_{\boldsymbol{k}}} \left(\frac{L}{N}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N-1} f\left(\frac{\boldsymbol{\ell} L}{N}\right) \mathrm{e}^{-\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{\ell}/N}, \qquad (*)$$

where

$$c_{\boldsymbol{k}} = c_{k_1} c_{k_2} c_{k_3}, \qquad c_{k_i} = \begin{cases} 1, & k_i = -N+1, \ldots, N-1, \\ 2, & k_i = \pm N. \end{cases}$$

The equation (*) is called the discrete Fourier transform of f.

Inverse DFT and fast Fourier transform (FFT)

The inversion of the DFT is given by

$$f\left(\frac{\boldsymbol{\ell} L}{N}\right) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{\ell}/N}, \qquad \boldsymbol{\ell} = -N, \ldots, N.$$

Inverse DFT and fast Fourier transform (FFT)

The inversion of the DFT is given by

$$f\left(\frac{\boldsymbol{\ell} L}{N}\right) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}} \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{\ell}/N}, \qquad \boldsymbol{\ell} = -N, \ldots, N.$$

Fast Fourier transform (FFT): The DFT and IDFT can be computed in $O(N^3 \log N)$ operations using the FFT algorithm!



- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues
- Quicksort Algorithm for Sorting
- Fast Fourier Transform
- Integer Relation Detection
- Fast Multipole Method

Application of FFT: convolution

Let $f(\boldsymbol{v}), g(\boldsymbol{v})$ be $2L$-periodic, and define

$$(f * g)(\boldsymbol{v}) := \int_{[-L,L]^3} f(\boldsymbol{v} - \boldsymbol{w}) g(\boldsymbol{w})\, d\boldsymbol{w}.$$

For full Fourier series (all modes $\boldsymbol{k} \in \mathbb{Z}^3$):

$$f(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k} \in \mathbb{Z}^3} \hat{f}_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k} \cdot \boldsymbol{v}/L}, \quad g(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k} \in \mathbb{Z}^3} \hat{g}_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k} \cdot \boldsymbol{v}/L}.$$

Then

$$\widehat{(f * g)}_{\boldsymbol{k}} = \hat{f}_{\boldsymbol{k}}\, \hat{g}_{\boldsymbol{k}}.$$

In computation we approximate f and g by f_N and g_N, containing only a finite number of modes $\boldsymbol{k} \in \{-N, \ldots, N\}^3$.

Application of FFT: convolution

Assume

$$f_N(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{f}_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{v}/L},$$

$$g_N(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{g}_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{v}/L}.$$

Define

$$h_N(\boldsymbol{v}) := (f_N * g_N)(\boldsymbol{v}).$$

Then

$$h_N(\boldsymbol{v}) = \frac{1}{(2L)^3} \sum_{\boldsymbol{k}=-N}^{N} \hat{h}_{\boldsymbol{k}}\, \mathrm{e}^{\mathrm{i}\pi \boldsymbol{k}\cdot\boldsymbol{v}/L}, \qquad \hat{h}_{\boldsymbol{k}} = \hat{f}_{\boldsymbol{k}}\, \hat{g}_{\boldsymbol{k}}.$$

FFT algorithm for convolution

FFT-based algorithm:

1. Compute $\hat{f}_{\boldsymbol{k}}$ and $\hat{g}_{\boldsymbol{k}}$ from sampled values by FFT using (*);
2. Form $\hat{h}_{\boldsymbol{k}} = \hat{f}_{\boldsymbol{k}} \hat{g}_{\boldsymbol{k}}$ for each $\boldsymbol{k} \in \{-N, \ldots, N\}^3$;
3. Recover h_N by inverse FFT if physical-space values are needed.

The coefficient multiplication costs $O(N^3)$, and the overall complexity including FFT and inverse FFT is $O(N^3 \log N)$.

1 Introduction to the Fourier spectral method

2 Fourier spectral method for the Boltzmann collision operator

3 Fix the physical properties

4 Generalization to spatially inhomogeneous Boltzmann equation

5 Summary

Review of the Boltzmann collision operator

Collision operator: given a distribution function $f(\boldsymbol{v})$,

$$Q(f,f)(\boldsymbol{v}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B\left(|\boldsymbol{v}-\boldsymbol{v}_1|, \frac{(\boldsymbol{v}-\boldsymbol{v}_1)\cdot\boldsymbol{\omega}}{|\boldsymbol{v}-\boldsymbol{v}_1|}\right) \times \left[f(\boldsymbol{v}')f(\boldsymbol{v}_1') - f(\boldsymbol{v})f(\boldsymbol{v}_1)\right] \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{v}_1,$$

where

$$\boldsymbol{v}' = \frac{\boldsymbol{v}+\boldsymbol{v}_1}{2} + \frac{|\boldsymbol{v}-\boldsymbol{v}_1|}{2}\boldsymbol{\omega}, \quad \boldsymbol{v}_1' = \frac{\boldsymbol{v}+\boldsymbol{v}_1}{2} - \frac{|\boldsymbol{v}-\boldsymbol{v}_1|}{2}\boldsymbol{\omega},$$

and B is the collision kernel.

Fourier transform of the collision operator

Fourier transform of $Q(f,f)$:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \underbrace{[K(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta}) - K(-\boldsymbol{\zeta},\boldsymbol{\zeta})]}_{\mathcal{C}(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta})} \hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta},$$

where

$$K(\boldsymbol{\eta},\boldsymbol{\zeta}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B\left(|\boldsymbol{g}|, \frac{\boldsymbol{g}\cdot\boldsymbol{\omega}}{|\boldsymbol{g}|}\right) \mathrm{e}^{-\frac{\mathrm{i}}{2}(\boldsymbol{\eta}-\boldsymbol{\zeta})\cdot\boldsymbol{g}} \mathrm{e}^{-\frac{\mathrm{i}}{2}|\boldsymbol{g}|(\boldsymbol{\eta}+\boldsymbol{\zeta})\cdot\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{g}.$$

Fourier transform of the collision operator

Fourier transform of $Q(f,f)$:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \underbrace{[K(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta}) - K(-\boldsymbol{\zeta},\boldsymbol{\zeta})]}_{\mathcal{C}(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta})} \hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta},$$

where

$$K(\boldsymbol{\eta},\boldsymbol{\zeta}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B\left(|\boldsymbol{g}|, \frac{\boldsymbol{g}\cdot\boldsymbol{\omega}}{|\boldsymbol{g}|}\right) \mathrm{e}^{-\frac{\mathrm{i}}{2}(\boldsymbol{\eta}-\boldsymbol{\zeta})\cdot\boldsymbol{g}} \mathrm{e}^{-\frac{\mathrm{i}}{2}|\boldsymbol{g}|(\boldsymbol{\eta}+\boldsymbol{\zeta})\cdot\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{g}.$$

Numerical approximation:

$$\hat{Q}_{\boldsymbol{k}} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}(\boldsymbol{\xi}_{\boldsymbol{k}}-\boldsymbol{\xi}_{\boldsymbol{\ell}},\boldsymbol{\xi}_{\boldsymbol{\ell}})\hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}}\hat{f}_{\boldsymbol{\ell}}.$$

Fourier transform of the collision operator

Fourier transform of $Q(f,f)$:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \underbrace{[K(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta}) - K(-\boldsymbol{\zeta},\boldsymbol{\zeta})]}_{\mathcal{C}(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta})} \hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta},$$

where

$$K(\boldsymbol{\eta},\boldsymbol{\zeta}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B\left(|\boldsymbol{g}|, \frac{\boldsymbol{g}\cdot\boldsymbol{\omega}}{|\boldsymbol{g}|}\right) \mathrm{e}^{-\frac{\mathrm{i}}{2}(\boldsymbol{\eta}-\boldsymbol{\zeta})\cdot\boldsymbol{g}} \mathrm{e}^{-\frac{\mathrm{i}}{2}|\boldsymbol{g}|(\boldsymbol{\eta}+\boldsymbol{\zeta})\cdot\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{g}.$$

Numerical approximation: $O(N^6)$

$$\hat{Q}_{\boldsymbol{k}} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}(\boldsymbol{\xi}_{\boldsymbol{k}}-\boldsymbol{\xi}_{\boldsymbol{\ell}},\boldsymbol{\xi}_{\boldsymbol{\ell}})\hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}}\hat{f}_{\boldsymbol{\ell}}.$$

Fourier spectral method for the collision operator

Given $f(\boldsymbol{v})$, we approximate $Q(f,f)(\boldsymbol{v})$ by

1. Approximate f by $\tilde{f}$ such that $\operatorname{supp}\tilde{f} \subset \{\boldsymbol{v} \mid |\boldsymbol{v}| \le R\}$.
2. Periodize $\tilde{f}$ to obtain a $2L$-periodic function $\tilde{f}_{\mathrm{per}}$, where $L \ge R$.
3. Compute $\hat{f}_{\boldsymbol{k}}$ from sampled values of $\tilde{f}_{\mathrm{per}}$ by FFT.
4. Compute $\hat{Q}_{\boldsymbol{k}}$ by

$$\hat{Q}_{\boldsymbol{k}} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}(\boldsymbol{\xi}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}, \boldsymbol{\xi}_{\boldsymbol{\ell}}) \hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}} \hat{f}_{\boldsymbol{\ell}}.$$

5. Recover the approximation of $Q(f,f)(\boldsymbol{v})$ by inverse FFT.

Fourier spectral method for the collision operator

Given $f(\boldsymbol{v})$, we approximate $Q(f,f)(\boldsymbol{v})$ by

1. Approximate f by $\tilde{f}$ such that $\operatorname{supp}\tilde{f} \subset \{\boldsymbol{v} \mid |\boldsymbol{v}| \le R\}$.
2. Periodize $\tilde{f}$ to obtain a $2L$-periodic function $\tilde{f}_{\mathrm{per}}$, where $L \ge R$.
3. Compute $\hat{f}_{\boldsymbol{k}}$ from sampled values of $\tilde{f}_{\mathrm{per}}$ by FFT.
4. Compute $\hat{Q}_{\boldsymbol{k}}$ by

$$\hat{Q}_{\boldsymbol{k}} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}(\boldsymbol{\xi}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}, \boldsymbol{\xi}_{\boldsymbol{\ell}}) \hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}} \hat{f}_{\boldsymbol{\ell}}.$$

5. Recover the approximation of $Q(f,f)(\boldsymbol{v})$ by inverse FFT.

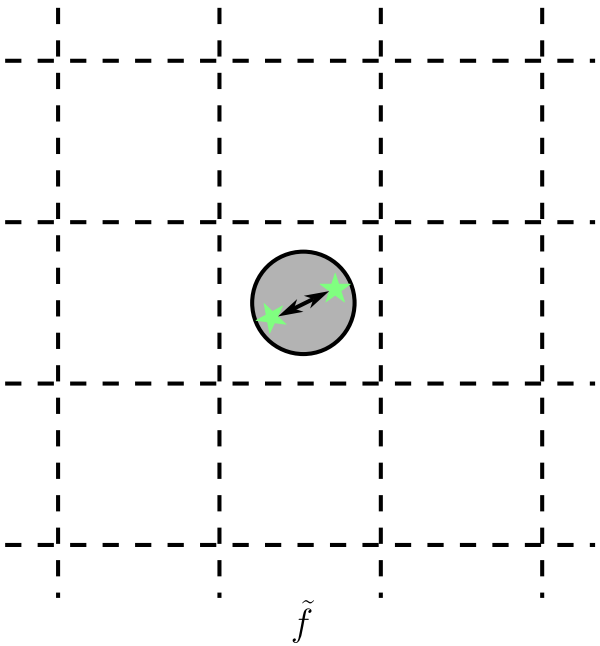
This method does not produce reliable approximations!

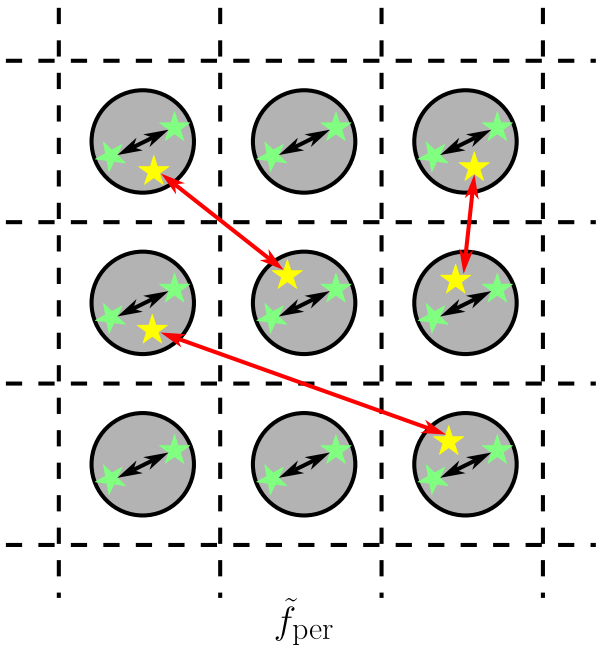
Aliasing error

For $\tilde{f}$ satisfying $\operatorname{supp} \tilde{f} \subset \{\boldsymbol{v} \mid |\boldsymbol{v}| \leq R\}$, we have

$$\operatorname{supp} Q(\tilde{f}, \tilde{f}) \subset \{\boldsymbol{v} \mid |\boldsymbol{v}| \leq \sqrt{2}R\}.$$

However, $\operatorname{supp} Q(\tilde{f}_{\mathrm{per}}, \tilde{f}_{\mathrm{per}})$ could be $\mathbb{R}^3$!





Dealiasing

Dealiasing:

- Eliminate all collisions with relative speed greater than $2R$.
- Choose a sufficiently large L such that particles with velocity in different periods do not collide.

Dealiasing

Dealiasing:

- Eliminate all collisions with relative speed greater than $2R$.
- Choose a sufficiently large L such that particles with velocity in different periods do not collide.

Collision term:

$$Q(f,f)(\boldsymbol{v}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B\left(|\boldsymbol{v}-\boldsymbol{v}_1|, \frac{(\boldsymbol{v}-\boldsymbol{v}_1)\cdot\boldsymbol{\omega}}{|\boldsymbol{v}-\boldsymbol{v}_1|}\right) \times \left[f(\boldsymbol{v}')f(\boldsymbol{v}_1') - f(\boldsymbol{v})f(\boldsymbol{v}_1)\right] \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{v}_1,$$

Dealiasing

Dealiasing:

- Eliminate all collisions with relative speed greater than $2R$.
- Choose a sufficiently large L such that particles with velocity in different periods do not collide.

Collision term:

$$Q^R(f,f)(\boldsymbol{v}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B^R\left(|\boldsymbol{v}-\boldsymbol{v}_1|, \frac{(\boldsymbol{v}-\boldsymbol{v}_1)\cdot\boldsymbol{\omega}}{|\boldsymbol{v}-\boldsymbol{v}_1|}\right) \times \left[f(\boldsymbol{v}')f(\boldsymbol{v}_1') - f(\boldsymbol{v})f(\boldsymbol{v}_1)\right] \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{v}_1,$$

where $B^R(g,\cos\theta) = B(g,\cos\theta)\mathbf{1}_{\{g\le 2R\}}$.

Fourier transform for the truncated operator

Fourier transform of $Q^R(f,f)$:

$$\widehat{Q^R(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \underbrace{[K^R(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta}) - K^R(-\boldsymbol{\zeta},\boldsymbol{\zeta})]}_{\mathcal{C}^R(\boldsymbol{\xi}-\boldsymbol{\zeta},\boldsymbol{\zeta})} \hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta},$$

where

$$K^R(\boldsymbol{\eta},\boldsymbol{\zeta}) = \int_{B(2R)}\int_{\mathbb{S}^2} B\left(|\boldsymbol{g}|, \frac{\boldsymbol{g}\cdot\boldsymbol{\omega}}{|\boldsymbol{g}|}\right) \mathrm{e}^{-\frac{\mathrm{i}}{2}(\boldsymbol{\eta}-\boldsymbol{\zeta})\cdot\boldsymbol{g}} \mathrm{e}^{-\frac{\mathrm{i}}{2}|\boldsymbol{g}|(\boldsymbol{\eta}+\boldsymbol{\zeta})\cdot\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{g}.$$

Fourier spectral method

Given $f(\boldsymbol{v})$, we approximate $Q^R(f,f)(\boldsymbol{v})$ by

1. Approximate f by $\tilde{f}$ such that $\operatorname{supp}\tilde{f} \subset \{\boldsymbol{v} \mid |\boldsymbol{v}| \le R\}$.
2. Periodize $\tilde{f}$ to obtain a $2L$-periodic function $\tilde{f}_{\mathrm{per}}$, where $L \ge 2R$.
3. Compute $\hat{f}_{\boldsymbol{k}}$ from sampled values of $\tilde{f}_{\mathrm{per}}$ by FFT.
4. Compute $\hat{Q}^R_{\boldsymbol{k}}$ by

$$\hat{Q}^R_{\boldsymbol{k}} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}^R(\boldsymbol{\xi}_{\boldsymbol{k}} - \boldsymbol{\xi}_{\boldsymbol{\ell}}, \boldsymbol{\xi}_{\boldsymbol{\ell}}) \hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}} \hat{f}_{\boldsymbol{\ell}}.$$

5. Recover the approximation of $Q^R(f,f)(\boldsymbol{v})$ by inverse FFT.

Spatially homogeneous Boltzmann equation

The spatially homogeneous Boltzmann equation is given by

$$\frac{\partial f}{\partial t} = Q(f, f), \qquad f(0, \boldsymbol{v}) = f_0(\boldsymbol{v}).$$

Spatially homogeneous Boltzmann equation

The spatially homogeneous Boltzmann equation is given by

$$\frac{\partial f}{\partial t} = Q(f,f), \qquad f(0,\boldsymbol{v}) = f_0(\boldsymbol{v}).$$

Fourier spectral method:

1. Approximate $f_0(\boldsymbol{v})$ by a $2L$-periodic function $\tilde{f}_{0,\mathrm{per}}$.
2. Compute $\hat{f}^0_{\boldsymbol{k}}$ from sampled values of $\tilde{f}_{0,\mathrm{per}}$ by FFT.
3. Solve the ODE system

$$\frac{\mathrm{d}\hat{f}_{\boldsymbol{k}}}{\mathrm{d}t} = \left(\frac{1}{2L}\right)^3 \sum_{\boldsymbol{\ell}=-N}^{N} \mathcal{C}^R(\boldsymbol{\xi}_{\boldsymbol{k}}-\boldsymbol{\xi}_{\boldsymbol{\ell}}, \boldsymbol{\xi}_{\boldsymbol{\ell}})\hat{f}_{\boldsymbol{k}-\boldsymbol{\ell}}\hat{f}_{\boldsymbol{\ell}}, \quad \hat{f}_{\boldsymbol{k}}(0) = \hat{f}^0_{\boldsymbol{k}}.$$

4. Recover the approximation of $f(t,\boldsymbol{v})$ by inverse FFT.

A fast algorithm for the collision operator

Another form of $\widehat{Q(f,f)}(\boldsymbol{\xi})$:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_0^{+\infty}\int_{\mathbb{S}^2} K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi})I(g,\boldsymbol{\sigma},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g - \frac{1}{2\pi^2}\int_{\mathbb{R}^3} K_-(\boldsymbol{\zeta})\hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta},$$

where

$$K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi}) = \int_{\mathbb{S}^2} g^2 B(g,\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{-\frac{\mathrm{i}}{2}g\boldsymbol{\xi}\cdot\boldsymbol{\omega}}\,\mathrm{d}\boldsymbol{\omega},$$

$$K_-(\boldsymbol{\zeta}) = \int_0^{+\infty}\int_{\mathbb{S}^2} \frac{g}{|\boldsymbol{\zeta}|}B(g,\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\sin\left(g|\boldsymbol{\zeta}|\right)\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g,$$

$$I(g,\boldsymbol{\sigma},\boldsymbol{\xi}) = \int_{\mathbb{R}^3} \hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\mathrm{e}^{\mathrm{i}g\boldsymbol{\sigma}\cdot(\boldsymbol{\xi}-\boldsymbol{\zeta})/2}\cdot\hat{f}(\boldsymbol{\zeta})\mathrm{e}^{-\mathrm{i}g\boldsymbol{\sigma}\cdot\boldsymbol{\zeta}/2}\,\mathrm{d}\boldsymbol{\zeta}.$$

A fast algorithm for the collision operator

Fourier transform of the collision operator:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_0^{+\infty}\int_{\mathbb{S}^2} K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi})I(g,\boldsymbol{\sigma},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g$$
$$- \frac{1}{2\pi^2}\int_{\mathbb{R}^3} K_-(\boldsymbol{\zeta})\hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta}.$$

A fast algorithm for the collision operator

Fourier transform of the collision operator:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_0^{+\infty}\int_{\mathbb{S}^2} K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi})I(g,\boldsymbol{\sigma},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g$$
$$- \frac{1}{2\pi^2}\int_{\mathbb{R}^3} K_-(\boldsymbol{\zeta})\hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta}.$$

Discretization:

- Discretize g by N points.
- Discretize $\boldsymbol{\sigma}$ by M points on the unit sphere.

A fast algorithm for the collision operator

Fourier transform of the collision operator:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_0^{+\infty}\int_{\mathbb{S}^2} K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi}) I(g,\boldsymbol{\sigma},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g - \frac{1}{2\pi^2}\int_{\mathbb{R}^3} K_-(\boldsymbol{\zeta})\hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta}.$$

Discretization:

- Discretize g by N points.
- Discretize $\boldsymbol{\sigma}$ by M points on the unit sphere.

Total computational cost: $O(MN^4 \log N)$.

A fast algorithm for the collision operator

Fourier transform of the collision operator:

$$\widehat{Q(f,f)}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^3}\int_0^{+\infty}\int_{\mathbb{S}^2} K_+(g,\boldsymbol{\sigma},\boldsymbol{\xi})I(g,\boldsymbol{\sigma},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\sigma}\,\mathrm{d}g - \frac{1}{2\pi^2}\int_{\mathbb{R}^3} K_-(\boldsymbol{\zeta})\hat{f}(\boldsymbol{\xi}-\boldsymbol{\zeta})\hat{f}(\boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta}.$$

Discretization:

- Discretize g by N points.
- Discretize $\boldsymbol{\sigma}$ by M points on the unit sphere.

Total computational cost: $O(MN^4 \log N)$.

Can you complete the details of the fast algorithm?

1 Introduction to the Fourier spectral method

2 Fourier spectral method for the Boltzmann collision operator

3 Fix the physical properties

4 Generalization to spatially inhomogeneous Boltzmann equation

5 Summary

Properties of the truncated collision operator

$$Q^R(f,f)(\boldsymbol{v}) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B^R\left(|\boldsymbol{v}-\boldsymbol{v}_1|, \frac{(\boldsymbol{v}-\boldsymbol{v}_1)\cdot\boldsymbol{\omega}}{|\boldsymbol{v}-\boldsymbol{v}_1|}\right) \times \left[f(\boldsymbol{v}')f(\boldsymbol{v}_1') - f(\boldsymbol{v})f(\boldsymbol{v}_1)\right] \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\boldsymbol{v}_1$$

- For any $2L$-periodic function f, $Q^R(f,f)$ is also $2L$-periodic.
- Mass conservation: $\displaystyle\int_{[-L,L]^3} Q^R(f,f)(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} = 0.$
- H-theorem: $\displaystyle\int_{[-L,L]^3} Q^R(f,f)(\boldsymbol{v})\log f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v} \le 0.$
- No momentum and energy conservation!
- The numerical scheme does not preserve the positivity!

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}^0_{\boldsymbol{k}}$ by FFT. Let $j = 0$.

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}^0_{\boldsymbol{k}}$ by FFT. Let $j = 0$.
2. Solve the ODE system for $\hat{f}_{\boldsymbol{k}}$ from time $j\Delta t$ to $(j+1)\Delta t$.

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}^0_{\boldsymbol{k}}$ by FFT. Let $j = 0$.
2. Solve the ODE system for $\hat{f}_{\boldsymbol{k}}$ from time $j\Delta t$ to $(j+1)\Delta t$.
3. Apply inv FFT to obtain $\tilde{f}^{j+1}_{\boldsymbol{\ell}}$, $\boldsymbol{\ell} \in \{-N, \cdots, N-1\}^3$, which approximate $f(t, \boldsymbol{\ell} L/N)$ at time $(j+1)\Delta t$.

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}_{\boldsymbol{k}}^0$ by FFT. Let $j = 0$.
2. Solve the ODE system for $\hat{f}_{\boldsymbol{k}}$ from time $j\Delta t$ to $(j+1)\Delta t$.
3. Apply inv FFT to obtain $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$, $\boldsymbol{\ell} \in \{-N, \cdots, N-1\}^3$, which approximate $f(t, \boldsymbol{\ell}L/N)$ at time $(j+1)\Delta t$.
4. Perturb $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$ to $f_{\boldsymbol{\ell}}^{j+1}$ such that $f_{\boldsymbol{\ell}}^{j+1} > 0$ and

$$\sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j+1} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j},$$

$$\sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j+1} \log f_{\boldsymbol{\ell}}^{j+1} \leq \sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j} \log f_{\boldsymbol{\ell}}^{j}.$$

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}_{\boldsymbol{k}}^0$ by FFT. Let $j = 0$.
2. Solve the ODE system for $\hat{f}_{\boldsymbol{k}}$ from time $j\Delta t$ to $(j+1)\Delta t$.
3. Apply inv FFT to obtain $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$, $\boldsymbol{\ell} \in \{-N, \cdots, N-1\}^3$, which approximate $f(t, \boldsymbol{\ell}L/N)$ at time $(j+1)\Delta t$.
4. Perturb $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$ to $f_{\boldsymbol{\ell}}^{j+1}$ such that $f_{\boldsymbol{\ell}}^{j+1} > 0$ and

$$\sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j+1} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j},$$

$$\sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j+1} \log f_{\boldsymbol{\ell}}^{j+1} \leq \sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j} \log f_{\boldsymbol{\ell}}^{j}.$$

5. Apply FFT to $f_{\boldsymbol{\ell}}^{j+1}$ to obtain $\hat{f}_{\boldsymbol{k}}^{j+1}$.

Fix the physical properties by postprocessing

1. Prepare initial Fourier coefficients $\hat{f}_{\boldsymbol{k}}^0$ by FFT. Let $j = 0$.
2. Solve the ODE system for $\hat{f}_{\boldsymbol{k}}$ from time $j\Delta t$ to $(j+1)\Delta t$.
3. Apply inv FFT to obtain $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$, $\boldsymbol{\ell} \in \{-N, \cdots, N-1\}^3$, which approximate $f(t, \boldsymbol{\ell}L/N)$ at time $(j+1)\Delta t$.
4. Perturb $\tilde{f}_{\boldsymbol{\ell}}^{j+1}$ to $f_{\boldsymbol{\ell}}^{j+1}$ such that $f_{\boldsymbol{\ell}}^{j+1} > 0$ and

$$\sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j+1} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j},$$

$$\sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j+1} \log f_{\boldsymbol{\ell}}^{j+1} \leq \sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j} \log f_{\boldsymbol{\ell}}^{j}.$$

5. Apply FFT to $f_{\boldsymbol{\ell}}^{j+1}$ to obtain $\hat{f}_{\boldsymbol{k}}^{j+1}$.
6. Let $j = j + 1$ and go to step 2.

Fix conservation and positivity by optimization

Idea: Minimize the distance between $\tilde{f}^{j+1}$ and f^{j+1}, subject to the constraints of conservation and positivity:

$$\min_{f_{\boldsymbol{k}}^{j+1}} \sum_{\boldsymbol{\ell}} |f_{\boldsymbol{\ell}}^{j+1} - \tilde{f}_{\boldsymbol{\ell}}^{j+1}|^2$$

$$\text{subject to} \quad \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j+1} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j}$$

$$f_{\boldsymbol{\ell}}^{j+1} \geqslant 0$$

Fix conservation and positivity by optimization

Idea: Minimize the distance between $\tilde{f}^{j+1}$ and f^{j+1}, subject to the constraints of conservation and positivity:

$$\min_{f_{\boldsymbol{k}}^{j+1}} \sum_{\boldsymbol{\ell}} |f_{\boldsymbol{\ell}}^{j+1} - \tilde{f}_{\boldsymbol{\ell}}^{j+1}|^2$$

$$\text{subject to} \quad \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j+1} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f_{\boldsymbol{\ell}}^{j}$$

$$f_{\boldsymbol{\ell}}^{j+1} \geqslant 0$$

Question: How to solve an optimization problem in the form

$$\min_{\boldsymbol{x}} \|\boldsymbol{x} - \boldsymbol{y}\|^2$$

$$\text{subject to} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{x} \geqslant 0?$$

Solving the optimization problem

Define

$$L(\boldsymbol{x};\boldsymbol{\lambda}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x}-\boldsymbol{b}) + \delta_+(\boldsymbol{x}),$$

where

$$\delta_+(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then

$$\max_{\boldsymbol{\lambda}} L(\boldsymbol{x};\boldsymbol{\lambda}) = \begin{cases} \|\boldsymbol{x}-\boldsymbol{y}\|^2, & \text{if } \mathbf{A}\boldsymbol{x}=\boldsymbol{b} \text{ and } \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Solving the optimization problem

Define

$$L(\boldsymbol{x};\boldsymbol{\lambda}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x}-\boldsymbol{b}) + \delta_+(\boldsymbol{x}),$$

where

$$\delta_+(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then

$$\max_{\boldsymbol{\lambda}} L(\boldsymbol{x};\boldsymbol{\lambda}) = \begin{cases} \|\boldsymbol{x}-\boldsymbol{y}\|^2, & \text{if } \mathbf{A}\boldsymbol{x}=\boldsymbol{b} \text{ and } \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

We just need to solve the unconstrained saddle point problem

$$\min_{\boldsymbol{x}} \max_{\boldsymbol{\lambda}} L(\boldsymbol{x};\boldsymbol{\lambda}).$$

Solving the optimization problem

Define

$$L(\boldsymbol{x};\boldsymbol{\lambda}) = \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x}-\boldsymbol{b}) + \delta_+(\boldsymbol{x}),$$

where

$$\delta_+(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then

$$\max_{\boldsymbol{\lambda}} L(\boldsymbol{x};\boldsymbol{\lambda}) = \begin{cases} \|\boldsymbol{x}-\boldsymbol{y}\|^2, & \text{if } \mathbf{A}\boldsymbol{x}=\boldsymbol{b} \text{ and } \boldsymbol{x} \geqslant 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

We just need to solve the unconstrained saddle point problem

$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}} L(\boldsymbol{x};\boldsymbol{\lambda}).$$

Solving the optimization problem

The inner minimization problem can be solved by

$$
\begin{aligned}
& \min_{\boldsymbol{x}} L(\boldsymbol{x};\boldsymbol{\lambda}) \\
= & \min_{\boldsymbol{x}} \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x}-\boldsymbol{b}) + \delta_+(\boldsymbol{x}) \\
= & \min_{\boldsymbol{x}\geqslant 0} \left\|\boldsymbol{x}-\left(\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right)\right\|^2 - \left\|\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right\|^2 + \|\boldsymbol{y}\|^2 - \boldsymbol{\lambda}^\top\boldsymbol{b} \\
= & -\left\|\max\left\{0,\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right\}\right\|^2 + \|\boldsymbol{y}\|^2 - \boldsymbol{\lambda}^\top\boldsymbol{b}.
\end{aligned}
$$

Solving the optimization problem

The inner minimization problem can be solved by

$$
\begin{aligned}
&\min_{\boldsymbol{x}} L(\boldsymbol{x};\boldsymbol{\lambda}) \\
=\ &\min_{\boldsymbol{x}} \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^\top(\mathbf{A}\boldsymbol{x}-\boldsymbol{b}) + \delta_+(\boldsymbol{x}) \\
=\ &\min_{\boldsymbol{x}\geqslant 0} \left\|\boldsymbol{x}-\left(\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right)\right\|^2 - \left\|\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right\|^2 + \|\boldsymbol{y}\|^2 - \boldsymbol{\lambda}^\top\boldsymbol{b} \\
=\ &-\left\|\max\left\{0,\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right\}\right\|^2 + \|\boldsymbol{y}\|^2 - \boldsymbol{\lambda}^\top\boldsymbol{b}.
\end{aligned}
$$

We just need to solve

$$
\max_{\boldsymbol{\lambda}} \Phi(\boldsymbol{\lambda}), \text{ with } \Phi(\boldsymbol{\lambda}) := -\left\|\max\left\{0,\boldsymbol{y}-\frac{1}{2}\mathbf{A}^\top\boldsymbol{\lambda}\right\}\right\|^2 - \boldsymbol{\lambda}^\top\boldsymbol{b}.
$$

Optimality condition for the dual problem

Define

$$\boldsymbol{x}(\boldsymbol{\lambda}) := \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}\right\}.$$

Maximizing Φ is equivalent to solving

$$\boldsymbol{F}(\boldsymbol{\lambda}) := \mathbf{A}\boldsymbol{x}(\boldsymbol{\lambda}) - \boldsymbol{b} = 0.$$

Since Φ is concave and piecewise quadratic, its maximizer $\boldsymbol{\lambda}^*$ satisfies

$$\mathbf{A}\max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}^*\right\} = \boldsymbol{b}.$$

Then recover the primal solution by

$$\boldsymbol{x}^* = \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}^*\right\}.$$

Semismooth Newton method

$$\boxed{\boldsymbol{F}(\boldsymbol{\lambda}) = 0}$$

Newton's method:

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \alpha_k \delta\boldsymbol{\lambda}^{(k)}, \qquad \mathbf{J}(\boldsymbol{\lambda}^{(k)})\delta\boldsymbol{\lambda}^{(k)} = -\boldsymbol{F}(\boldsymbol{\lambda}^{(k)}).$$

where $\mathbf{J}(\boldsymbol{\lambda}^{(k)}) = \nabla_\lambda \boldsymbol{F}(\boldsymbol{\lambda}^{(k)})$.

Semismooth Newton method

$$\boxed{\boldsymbol{F}(\boldsymbol{\lambda}) = 0}$$

Newton's method:

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \alpha_k \delta\boldsymbol{\lambda}^{(k)}, \qquad \mathbf{J}(\boldsymbol{\lambda}^{(k)})\delta\boldsymbol{\lambda}^{(k)} = -\boldsymbol{F}(\boldsymbol{\lambda}^{(k)}).$$

where $\mathbf{J}(\boldsymbol{\lambda}^{(k)}) = \nabla_\lambda \boldsymbol{F}(\boldsymbol{\lambda}^{(k)})$.

However, $\boldsymbol{F}$ is not differentiable for some $\boldsymbol{\lambda}$.

Semismooth Newton method

$$\boxed{\boldsymbol{F}(\boldsymbol{\lambda}) = 0}$$

Semismooth Newton method:

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \alpha_k \delta\boldsymbol{\lambda}^{(k)}, \qquad \mathbf{J}(\boldsymbol{\lambda}^{(k)})\delta\boldsymbol{\lambda}^{(k)} = -\boldsymbol{F}(\boldsymbol{\lambda}^{(k)}).$$

where $\mathbf{J}(\boldsymbol{\lambda}^{(k)}) \in \partial\boldsymbol{F}(\boldsymbol{\lambda}^{(k)})$.

Semismooth Newton method

$$\boxed{\boldsymbol{F}(\boldsymbol{\lambda}) = 0}$$

Semismooth Newton method:

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \alpha_k \delta\boldsymbol{\lambda}^{(k)}, \qquad \mathbf{J}(\boldsymbol{\lambda}^{(k)})\delta\boldsymbol{\lambda}^{(k)} = -\boldsymbol{F}(\boldsymbol{\lambda}^{(k)}).$$

where $\mathbf{J}(\boldsymbol{\lambda}^{(k)}) \in \partial\boldsymbol{F}(\boldsymbol{\lambda}^{(k)})$.

Generalized Jacobian:

$$\partial\boldsymbol{F}(\boldsymbol{\lambda}) := \operatorname{conv}\left\{\lim_{k\to\infty} \nabla\boldsymbol{F}(\boldsymbol{\lambda}^{(k)}) \mid \boldsymbol{\lambda}^{(k)} \to \boldsymbol{\lambda},\ \boldsymbol{\lambda}^{(k)} \notin S_{\boldsymbol{F}}\right\}.$$

Here $S_{\boldsymbol{F}}$ is the set of non-differentiability points.

Generalized Jacobian for this problem

For

$$\boldsymbol{F}(\boldsymbol{\lambda}) = \mathbf{A} \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}\right\} - \boldsymbol{b},$$

its generalized Jacobian is

$$\mathbf{J}(\boldsymbol{\lambda}) = -\frac{1}{2}\mathbf{A}\mathbf{D}(\boldsymbol{\lambda})\mathbf{A}^\top,$$

where $\mathbf{D}(\boldsymbol{\lambda}) = \operatorname{diag}(d_1, \ldots, d_n)$ with

$$d_i = \begin{cases} 1, & y_i - \frac{1}{2}(\mathbf{A}^\top \boldsymbol{\lambda})_i > 0, \\ 0, & y_i - \frac{1}{2}(\mathbf{A}^\top \boldsymbol{\lambda})_i \leqslant 0. \end{cases}$$

Generalized Jacobian for this problem

For

$$\boldsymbol{F}(\boldsymbol{\lambda}) = \mathbf{A} \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}\right\} - \boldsymbol{b},$$

its generalized Jacobian is

$$\mathbf{J}(\boldsymbol{\lambda}) = -\frac{1}{2}\mathbf{A}\mathbf{D}(\boldsymbol{\lambda})\mathbf{A}^\top,$$

where $\mathbf{D}(\boldsymbol{\lambda}) = \mathrm{diag}(d_1, \ldots, d_n)$ with

$$d_i = \begin{cases} 1, & y_i - \frac{1}{2}(\mathbf{A}^\top \boldsymbol{\lambda})_i > 0, \\ 0, & y_i - \frac{1}{2}(\mathbf{A}^\top \boldsymbol{\lambda})_i \leqslant 0. \end{cases}$$

For our application, $\mathbf{A}$ is a $5 \times (2N)^3$ matrix, and $\mathbf{J}(\boldsymbol{\lambda})$ is a 5×5 matrix.

Algorithm for fixing conservation and positivity

Algorithm 1 Algorithm for fixing conservation and positivity

Input: Vector $\boldsymbol{y}$, matrix $\mathbf{A}$, vector $\boldsymbol{b}$, tolerance ε
Choose an initial multiplier $\boldsymbol{\lambda}^{(0)}$ and set $k \leftarrow 0$
$\boldsymbol{x}^{(k)} = \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}^{(k)}\right\}$, $\boldsymbol{r}^{(k)} = \mathbf{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}$
while $\|\boldsymbol{r}^{(k)}\| > \varepsilon$ **do**
    Form $\mathbf{D}_k = \mathrm{diag}(d_i^{(k)})$ with $d_i^{(k)} = \mathbf{1}_{\{x_i^{(k)}>0\}}$
    Solve $\mathbf{A}\mathbf{D}_k\mathbf{A}^\top \delta\boldsymbol{\lambda}^{(k)} = 2\boldsymbol{r}^{(k)}$
    Choose $\alpha_k \in (0, 1]$ and update $\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \alpha_k \delta\boldsymbol{\lambda}^{(k)}$
    Set $k \leftarrow k + 1$ and compute:

$$\boldsymbol{x}^{(k)} = \max\left\{0, \boldsymbol{y} - \frac{1}{2}\mathbf{A}^\top \boldsymbol{\lambda}^{(k)}\right\}, \quad \boldsymbol{r}^{(k)} = \mathbf{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}$$

end while
Output: $\boldsymbol{x}^* = \boldsymbol{x}^{(k)}$

Fix the entropy inequality

We now have $f_\ell^{j+1} \geq 0$ that conserves mass, momentum, and energy. How can we ensure the entropy inequality

$$\sum_\ell f_\ell^{j+1} \log f_\ell^{j+1} \leq \sum_\ell f_\ell^j \log f_\ell^j ? \qquad (*)$$

Fix the entropy inequality

We now have $f_\ell^{j+1} \geq 0$ that conserves mass, momentum, and energy. How can we ensure the entropy inequality

$$\sum_\ell f_\ell^{j+1} \log f_\ell^{j+1} \leq \sum_\ell f_\ell^j \log f_\ell^j ? \qquad (*)$$

1. If (*) holds, do nothing.
2. If (*) is violated, find $\sigma \in [0, 1)$ such that

$$\sum_\ell [\underbrace{\sigma f_\ell^{j+1} + (1-\sigma)\mathcal{M}_\ell^j}_{\text{updated solution}}] \log[\sigma f_\ell^{j+1} + (1-\sigma)\mathcal{M}_\ell^j] = \sum_\ell f_\ell^j \log f_\ell^j,$$

where $\mathcal{M}^j$ is the maximum entropy distribution function with the same mass, momentum, and energy as f^j.

The maximum entropy distribution

Solve for $\mathcal{M}^j$:

$$\min_{\mathcal{M}^j} \sum_{\boldsymbol{\ell}} \mathcal{M}^j_{\boldsymbol{\ell}} \log \mathcal{M}^j_{\boldsymbol{\ell}}$$

$$\text{subject to} \quad \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) \mathcal{M}^j_{\boldsymbol{\ell}} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f^j_{\boldsymbol{\ell}}.$$

The maximum entropy distribution

Solve for $\mathcal{M}^j$:

$$\min_{\mathcal{M}^j} \sum_{\boldsymbol{\ell}} \mathcal{M}^j_{\boldsymbol{\ell}} \log \mathcal{M}^j_{\boldsymbol{\ell}}$$

$$\text{subject to} \quad \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) \mathcal{M}^j_{\boldsymbol{\ell}} = \sum_{\boldsymbol{\ell}} (1, \boldsymbol{\ell}, |\boldsymbol{\ell}|^2) f^j_{\boldsymbol{\ell}}.$$

Solution:

$$\mathcal{M}^j_{\boldsymbol{\ell}} = \exp\left(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{\ell} + \gamma |\boldsymbol{\ell}|^2\right),$$

where $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^3$, and $\gamma \in \mathbb{R}$ are Lagrange multipliers determined by the constraints, which can be solved by Newton's method.

Algorithm for fixing the entropy inequality

Algorithm 2 Fix the entropy inequality using bisection

Input: $f_{\boldsymbol{\ell}}^{j+1}$, $f_{\boldsymbol{\ell}}^{j}$, ε
Compute $H^{j+1} = \sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j+1} \log f_{\boldsymbol{\ell}}^{j+1}$ and $H^{j} = \sum_{\boldsymbol{\ell}} f_{\boldsymbol{\ell}}^{j} \log f_{\boldsymbol{\ell}}^{j}$
If $H^{j+1} \leqslant H^{j}$, output $f_{\boldsymbol{\ell}}^{j+1}$ and stop
Compute $\mathcal{M}_{\boldsymbol{\ell}}^{j} = \exp(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{\ell} + \gamma |\boldsymbol{\ell}|^2)$ with the same moments as f^{j}
Set $\sigma_L \leftarrow 0$, $\sigma_R \leftarrow 1$, and $g_{\boldsymbol{\ell}}(\sigma) = \sigma f_{\boldsymbol{\ell}}^{j+1} + (1-\sigma)\mathcal{M}_{\boldsymbol{\ell}}^{j}$
while $\sigma_R - \sigma_L > \varepsilon$ **do**
    $\sigma \leftarrow (\sigma_L + \sigma_R)/2$
    If $\sum_{\boldsymbol{\ell}} g_{\boldsymbol{\ell}}(\sigma) \log g_{\boldsymbol{\ell}}(\sigma) > H^{j}$, set $\sigma_R \leftarrow \sigma$; otherwise set $\sigma_L \leftarrow \sigma$
end while
Output: $f_{\boldsymbol{\ell}}^{j+1} \leftarrow \sigma_L f_{\boldsymbol{\ell}}^{j+1} + (1-\sigma_L)\mathcal{M}_{\boldsymbol{\ell}}^{j}$

1 Introduction to the Fourier spectral method

2 Fourier spectral method for the Boltzmann collision operator

3 Fix the physical properties

4 Generalization to spatially inhomogeneous Boltzmann equation

5 Summary

Time-dependent Boltzmann equation

Boltzmann equation:

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f = Q(f, f), \qquad f(0, \boldsymbol{x}, \boldsymbol{v}) = f_0(\boldsymbol{x}, \boldsymbol{v}).$$

Time-dependent Boltzmann equation

Boltzmann equation:

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f = Q(f, f), \qquad f(0, \boldsymbol{x}, \boldsymbol{v}) = f_0(\boldsymbol{x}, \boldsymbol{v}).$$

Operator splitting: One time step from $t_j = j\Delta t$ to $t_{j+1} = (j+1)\Delta t$ consists of two steps:

1. Solve the transport step

$$\partial_t f + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f = 0$$

from t_j to t_{j+1} on a spatial grid for every velocity $\boldsymbol{v}_{\boldsymbol{\ell}} = \boldsymbol{\ell} L/N$.

2. Solve the collision step

$$\partial_t f = Q(f, f)$$

from t_j to t_{j+1} for every spatial grid point.

Time-independent (steady-state) equation

Steady-state Boltzmann equation:

$$\boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f = Q(f, f)$$

with some boundary conditions.

Time-independent (steady-state) equation

Steady-state Boltzmann equation:

$$\boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f = Q(f, f)$$

with some boundary conditions.

Fixed-point iteration:

$$\boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f^{(k+1)} + \nu^{(k)} f^{(k+1)} = Q(f^{(k)}, f^{(k)}) + \nu^{(k)} f^{(k)},$$

where $\nu^{(k)}(\boldsymbol{x})$ is chosen such that the right-hand side is nonnegative.

In the next lecture

- The fixed-point iteration is slow when the gas is dense.
- A new iteration scheme will be introduced to accelerate the convergence.

1 Introduction to the Fourier spectral method

2 Fourier spectral method for the Boltzmann collision operator

3 Fix the physical properties

4 Generalization to spatially inhomogeneous Boltzmann equation

5 Summary

Summary

Fourier spectral method:

- Discretization for functions on unbounded domains
- Dealiasing for the collision operator

Structure-preserving post-processing:

- Fix conservation and positivity by solving an optimization problem
- Fix the entropy inequality by a convex combination with the maximum entropy distribution