4. Some statistical mechanics

4.1 The Boltzmann equation

The motion of a fluid can be described on various levels. The most basic decription by the Hamilton equations for a set of classical particles or the analogous quantum mechanical formulation prohibits itself because of the huge number of particles. 1 cm^3 of air, for example, at 0°C and a pressure of one atmosphere contains $2.69 \cdot 10^{19}$ molecules. It is impossible to prepare a desired microstate of such a system.
This fact is already taken into account by the next higher level of decription for a system with N particles by distribution functions $f_N(\boldsymbol{q}_1, \boldsymbol{p}_1, ..., \boldsymbol{q}_N, \boldsymbol{p}_N, t)$ which encompass all statistical informations of all dynamical processes ($\boldsymbol{q}_i$ and $\boldsymbol{p}_i$ are the generalized coordinate and momentum of particle i). $f_N(\boldsymbol{q}_1, \boldsymbol{p}_1, ..., \boldsymbol{q}_N, \boldsymbol{p}_N, t) d\boldsymbol{q}_1 d\boldsymbol{p}_1 ... d\boldsymbol{q}_N d\boldsymbol{p}_N$ is the probability to find a particle in the interval $([\boldsymbol{q}_1, \boldsymbol{q}_1 + d\boldsymbol{q}_1], [\boldsymbol{p}_1, \boldsymbol{p}_1 + d\boldsymbol{p}_1])$ while the other particles are in infinitesimal intervals around $(\boldsymbol{q}_2, \boldsymbol{p}_2)$ ... $(\boldsymbol{q}_N, \boldsymbol{p}_N)$. Thus f_N contains especially the various correlations between particles. f_N obeys the *Liouville equation*

$$\frac{\partial f_N}{\partial t} - \sum_{j=1}^{3N} \left(\frac{\partial H_N}{\partial q_j} \frac{\partial f_N}{\partial p_j} - \frac{\partial H_N}{\partial p_j} \frac{\partial f_N}{\partial q_j} \right) = 0 \tag{4.1.1}$$

where H_N is the Hamiltonian of the system.
By integration over part of the phase space one defines *reduced densities*

$$F_s(\boldsymbol{q}_1, \boldsymbol{p}_1, ..., \boldsymbol{q}_s, \boldsymbol{p}_s, t) := V^s \int f_N(\boldsymbol{q}_1, \boldsymbol{p}_1, ..., \boldsymbol{q}_N, \boldsymbol{p}_N, t)\, d\boldsymbol{q}_{s+1} d\boldsymbol{p}_{s+1} ... d\boldsymbol{q}_N d\boldsymbol{p}_N$$

where V^s is a normalization factor. It has been shown that a coupled system of differential equations for the F_s $(1 \leq s \leq N)$ is equivalent to the Liouville equation. This system is called *BBGKY* after Bogoljubov, Born, Green, Kirkwood and Yvon who derived these equations. The BBGKY hierarchy has to be truncated at some point to calculate approximate solutions.
The *Boltzmann equation* has been derived as a result of a systematic approximation starting from the BBGKY system not before 1946 (compare Bogoliubov, 1962; Boltzmann derived the equation which bears his name by a

D.A. Wolf-Gladrow: LNM 1725, pp. 139–158, 2000.

different reasoning already in the 19th century). It can be derived by applying the following approximations: 1. Only two-particle collisions are considered (this seems to restict applications to dilute gases). 2. The velocities of the two colliding particles are uncorrelated before collision. This assumption is often called the *molecular chaos hypothesis*. 3. External forces do not influence the local collision dynamics.
The Boltzmann equation is an integro-differential equation for the single particle distribution function $f(\boldsymbol{x}, \boldsymbol{v}, t) \propto F_1(\boldsymbol{q}_1, \boldsymbol{p}_1, t)$

$$\partial_t f + \boldsymbol{v}\partial_{\boldsymbol{x}} f + \frac{\boldsymbol{K}}{m}\partial_{\boldsymbol{v}} f = Q(f, f) \tag{4.1.2}$$

where $\boldsymbol{x} = \boldsymbol{q}_1$, $\boldsymbol{v} = \boldsymbol{p}_1/m$, $m = const$ is the particle mass, $f(\boldsymbol{x}, \boldsymbol{v}, t)\, d^3x\, d^3v$ is the probability to find a particle in the volume d^3x around $\boldsymbol{x}$ and with velocity between $\boldsymbol{v}$ and $\boldsymbol{v} + d\boldsymbol{v}$.

$$Q(f, f) = \int d^3v_1 \int d\Omega\, \sigma(\Omega)|\boldsymbol{v} - \boldsymbol{v}_1|[f(\boldsymbol{v}')f(\boldsymbol{v}_1') - f(\boldsymbol{v})f(\boldsymbol{v}_1)] \tag{4.1.3}$$

is the *collision integral* with $\sigma(\Omega)$ the *differential collision cross section* for the two-particle collision which transforms the velocities from $\{\boldsymbol{v}, \boldsymbol{v}_1\}$ (incoming) into $\{\boldsymbol{v}', \boldsymbol{v}_1'\}$ (outgoing). $\boldsymbol{K}$ is the body force. It will be neglected in the following discussion of the current chapter.

4.1.1 Five collision invariants and Maxwell's distribution

It can be shown (see, for example, Cercignani, 1988) that the collision integral possesses exactly five elementary *collision invariants* $\psi_k(\boldsymbol{v})$ $(k = 0, 1, 2, 3, 4)$ in the sense that

$$\int Q(f, f)\psi_k(\boldsymbol{v})\, d^3v = 0. \tag{4.1.4}$$

The elementary collision invariants read $\psi_0 = 1$, $(\psi_1, \psi_2, \psi_3) = \boldsymbol{v}$ and $\psi_4 = \boldsymbol{v}^2$ (proportional to mass, momentum and kinetic energy). General collision invariants $\phi(\boldsymbol{v})$ can be written as linear combinations of the ψ_k

$$\phi(\boldsymbol{v}) = a + \boldsymbol{b} \cdot \boldsymbol{v} + c\boldsymbol{v}^2.$$

It can be further shown (see, for example, Cercignani, 1988) that positive functions f exist which give a vanishing collision integral

$$Q(f, f) = 0.$$

These functions are all of the form

$$f(\boldsymbol{v}) = exp(a + \boldsymbol{b} \cdot \boldsymbol{v} + c\boldsymbol{v}^2)$$

where c must be negative. The *Maxwell*[1] *distribution*

$$\boxed{f^{(M)} = f(\boldsymbol{x}, \boldsymbol{v}, t) = n \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left[-\frac{m}{2k_B T} (\boldsymbol{v} - \boldsymbol{u})^2 \right]} \tag{4.1.5}$$

is a special case among these solutions where $\boldsymbol{u}$ is the mean velocity

$$\boldsymbol{u} = \frac{1}{n} \int d^3v \, \boldsymbol{v} f(\boldsymbol{x}, \boldsymbol{v}, t). \tag{4.1.6}$$

Please note that $f^{(M)}$ depends on $\boldsymbol{x}$ only implicitly via $n(\boldsymbol{x})$, $\boldsymbol{u}(\boldsymbol{x})$ and $T(\boldsymbol{x})$.

4.1.2 Boltzmann's H-theorem

In 1872 Boltzmann showed that the quantity

$$H(t) := \int d^3v \, d^3x \, f(\boldsymbol{x}, \boldsymbol{v}, t) \ln f(\boldsymbol{x}, \boldsymbol{v}, t) \tag{4.1.7}$$

where $f(\boldsymbol{x}, \boldsymbol{v}, t)$ is any function that satisfies the Boltzmann equation fulfills the equation

$$\boxed{\frac{dH}{dt} \leq 0} \tag{4.1.8}$$

and the equal sign applies only if f is a Maxwell distribution (4.1.5). This is his famous *H-theorem*.

Proof. We will assume that no external forces are applied and thus $f(\boldsymbol{x}, \boldsymbol{v}, t)$ obeys the following Botzmann equation

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \boldsymbol{\nabla} f = Q(f, f).$$

Differentiation of (4.1.7) yields

$$\frac{dH}{dt} = \int d^3v \, d^3x \, \frac{\partial f}{\partial t} (\boldsymbol{x}, \boldsymbol{v}, t) \left[1 + \ln f(\boldsymbol{x}, \boldsymbol{v}, t) \right].$$

Insertion of the Boltzmann equation leads to

$$\begin{aligned} \frac{dH}{dt} &= -\int \boldsymbol{v} \cdot \boldsymbol{\nabla} \left[f(\boldsymbol{x}, \boldsymbol{v}, t) \ln f(\boldsymbol{x}, \boldsymbol{v}, t) \right] d^3v \, d^3x \\ &+ \int d^3v_1 \, d^3v_2 \, d^3x \, d\Omega \, \sigma(\Omega) \left[f_2' f_1' - f_2 f_1 \right] \cdot \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid \left[1 + \ln f_1 \right] \end{aligned} \tag{4.1.9}$$

[1] The *Maxwell distribution* is also referred to as *Maxwell-Boltzmann distribution* or as *Boltzmann distribution*.

where $f_1 = f(\boldsymbol{x}, \boldsymbol{v}_1, t)$, $f_2 = f(\boldsymbol{x}, \boldsymbol{v}_2, t)$, $f_1' = f(\boldsymbol{x}, \boldsymbol{v}_1', t)$, and $f_2' = f(\boldsymbol{x}, \boldsymbol{v}_2', t)$. The first summand can be transformed into a surface integral

$$-\int_F \boldsymbol{n} \cdot \boldsymbol{v} f(\boldsymbol{x}, \boldsymbol{v}, t) \ln f(\boldsymbol{x}, \boldsymbol{v}, t)\, d^3v\, dF \tag{4.1.10}$$

where $\boldsymbol{n}$ is the (outer) normal of the surface F that enclosed the gas. Without detailed discussion we will assume that this surface integral vanishes. The second integral is invariant under exchange of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ because $\sigma(\Omega)$ is invariant under such exchange:

$$\frac{dH}{dt} = \int d^3v_1\, d^3v_2\, d^3x\, d\Omega\, \sigma(\Omega) \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid (f_2' f_1' - f_2 f_1)[1 + \ln f_2] \tag{4.1.11}$$

Adding up half of (4.1.9) and half of (4.1.11) leads to

$$\frac{dH}{dt} = \frac{1}{2}\int d^3v_1\, d^3v_2\, d^3x\, d\Omega\, \sigma(\Omega) \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid (f_2' f_1' - f_2 f_1)[2 + \ln(f_1 f_2)] \tag{4.1.12}$$

This integral is invariant under exchange of $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ and $\{\boldsymbol{v}_1', \boldsymbol{v}_2'\}$ because for each collision there exists an inverse collision with the same cross section. Therefore, one obtains

$$\frac{dH}{dt} = \frac{1}{2}\int d^3v_1'\, d^3v_2'\, d^3x\, d\Omega\, \sigma'(\Omega) \mid \boldsymbol{v}_2' - \boldsymbol{v}_1' \mid (f_2 f_1 - f_2' f_1')[2 + \ln(f_1' f_2')]$$

and because of $d^3v_1' d^3v_2' = d^3v_1 d^3v_2$, $\mid \boldsymbol{v}_2' - \boldsymbol{v}_1' \mid = \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid$ and $\sigma'(\Omega) = \sigma(\Omega)$:

$$\frac{dH}{dt} = \frac{1}{2}\int d^3v_1\, d^3v_2\, d^3x\, d\Omega\, \sigma(\Omega) \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid (f_2 f_1 - f_2' f_1')[2 + \ln(f_1' f_2')]. \tag{4.1.13}$$

Adding up half of (4.1.12) and half of (4.1.13) leads to

$$\frac{dH}{dt} = \frac{1}{2}\int d^3v_1\, d^3v_2\, d^3x\, d\Omega\, \sigma(\Omega) \mid \boldsymbol{v}_2 - \boldsymbol{v}_1 \mid (f_2' f_1' - f_2 f_1)[\ln(f_1 f_2) - \ln(f_1' f_2')].$$

The integrand is never positive because of the inequality

$$(b - a) \cdot (\ln a - \ln b) > 0, \quad a \neq b > 0,$$

thus $dH/dt \leq 0$.

It vanishes, however, when $(f_2' f_1' - f_2 f_1) = 0$ and therefore $\dfrac{\partial f(\boldsymbol{v}, t)}{\partial t} = 0$. $dH/dt = 0$ is possible if and only if

$$f(\boldsymbol{v}_1') f(\boldsymbol{v}_2') - f(\boldsymbol{v}_1) f(\boldsymbol{v}_2) = 0 \tag{4.1.14}$$

for all $\boldsymbol{v}_1'$, $\boldsymbol{v}_2'$ that result from $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ by collisions. From (4.1.14) one obtains

$$\ln f(\boldsymbol{v}_1') + \ln f(\boldsymbol{v}_2') = \ln f(\boldsymbol{v}_1) + \ln f(\boldsymbol{v}_2), \tag{4.1.15}$$

i.e. $\ln f(\boldsymbol{v})$ is an additive collision invariant and thus it is of the form (linear composition of the five collision invariants):

$$\ln f(\boldsymbol{x}, \boldsymbol{v}) = a(\boldsymbol{x}) + \boldsymbol{b}(\boldsymbol{x}) \cdot \boldsymbol{v} + c(\boldsymbol{x})\boldsymbol{v}^2. \tag{4.1.16}$$

Therefore it follows that

$$f(\boldsymbol{x}, \boldsymbol{v}) = C(\boldsymbol{x}) e^{-\frac{m(\boldsymbol{v} - \boldsymbol{u}(\boldsymbol{x}))^2}{2k_B T(\boldsymbol{x})}} \tag{4.1.17}$$

where $C(\boldsymbol{x})$, $u(\boldsymbol{x})$ and $T(\boldsymbol{x})$ are independent of $\boldsymbol{v}$. However, the distribution (4.1.17) represents no equilibrium state because if $f(\boldsymbol{x}, \boldsymbol{v}, t_1)$ at time t_1 is of the form (4.1.17) then it follows from the Boltzmann equation that

$$\left(\frac{\partial f(\boldsymbol{x}, \boldsymbol{v}, t)}{\partial t}\right)_{t=t_1} = -\boldsymbol{v} \cdot \boldsymbol{\nabla}_{\boldsymbol{x}} \left[C(\boldsymbol{x}) e^{-\frac{m(\boldsymbol{v} - \boldsymbol{u}(\boldsymbol{x}))^2}{2T(\boldsymbol{x})}} \right] \tag{4.1.18}$$

(the collision term $Q(f, f)$ vanishes because f($\boldsymbol{x}$,$\boldsymbol{v}$) is a function of collision invariants). For the equilibrium state f must be of the form (4.1.17) and be independent of $\boldsymbol{x}$ thus

$$f^{(eq)}(\boldsymbol{x}, \boldsymbol{v}) = f^{(eq)}(\boldsymbol{v}) = C e^{-\frac{m(\boldsymbol{v} - \boldsymbol{u})^2}{2k_B T}} \tag{4.1.19}$$

with *constants* C, T, $\boldsymbol{u}$. In a closed system at rest the mean velocity $\boldsymbol{u}$ must vanish and therefore

$$f^{(eq)}(\boldsymbol{v}) = C e^{-\frac{m\boldsymbol{v}^2}{2k_B T}}. \tag{4.1.20}$$

This is the famous Maxwell velocity distribution. **q.e.d.**

4.1.3 The BGK approximation

One of the major problems when dealing with the Boltzmann equation is the complicated nature of the collision integral. It is therefore not surprising that alternative, simpler expressions have been proposed. The idea behind this replacement is that the large amount of detail of two-body interactions is not likely to influence significantly the values of many experimentally measured quantities (Cercignani, 1990).
The simpler operator $J(f)$ which replaces the collision operator $Q(f, f)$ should respect two constraints:

1. $J(f)$ conserves the collision invariants ψ_k of $Q(f, f)$, that is

$$\int \psi_k J(f)\, d^3x\, d^3v = 0 \quad (k = 0, 1, 2, 3, 4), \tag{4.1.21}$$

2. The collision term expresses the tendency to a Maxwellian distribution (H-theorem).

Both constraints are fulfilled by the most widely known model called usually the BGK approximation. It was proposed by Bhatnagar, Gross and Krook (1954) and independently at about the same time by Welander (1954). The simplest way to take the second constraint into account is to imagine that each collision changes the distribution function $f(\boldsymbol{x}, \boldsymbol{v})$ by an amount proportional to the departure of f from a Maxwellian $f^M(\boldsymbol{x}, \boldsymbol{v})$:

$$J(f) = \omega \left[f^M(\boldsymbol{x}, \boldsymbol{v}) - f(\boldsymbol{x}, \boldsymbol{v}) \right]. \tag{4.1.22}$$

The coefficient ω is called the collision frequency. From the first constraint it follows

$$\int \psi_k J(f)\, d^3x\, d^3v = \omega \left[\int \psi_k f^M(\boldsymbol{x}, \boldsymbol{v})\, d^3x\, d^3v - \int \psi_k f(\boldsymbol{x}, \boldsymbol{v})\, d^3x\, d^3v \right] = 0 \tag{4.1.23}$$

i.e. at any space point and time instant the Maxwellian $f^M(\boldsymbol{x}, \boldsymbol{v})$ must have exactly the same density, velocity and temperature of the gas as given by the distribution $f(\boldsymbol{x}, \boldsymbol{v})$. Since these values will in general vary with space and time $f^M(\boldsymbol{x}, \boldsymbol{v})$ is called the *local Maxwellian*. Other *model equations* are discussed in Cercignani (1990).

4.2 Chapman-Enskog: From Boltzmann to Navier-Stokes

Many fluid-dynamical phenomena including laminar flows, turbulence and solitons can be described by solutions of the Navier-Stokes equation. Although the form of this equation can be obtained by phenomenological reasoning (see, for example, Landau and Lifshitz, 1959) it is of fundamental as well as practical interest to derive the Navier-Stokes equation (Eq. 1.3.1) from the Boltzmann equation. Applying certain models of the microscopic collision processes one can obtain explicit formulas for the transport coefficients. For example, Maxwell was able to derive an analytical expression for the shear viscosity for molecules which interact by a r^{-5}-potential where r is their distance. It came as a surprise for him and his contemporaries that this theory predicted a dynamic viscosity coefficient independent of density. Experiments made thereafter indeed showed that this is a good approximation for gases over a wide range of densities.

The derivation of the Navier-Stokes equation and its transport coefficients from the Boltzmann equation and certain microscopic collision models runs under the name Chapman-Enskog expansion. This method has been developed by Chapman and Enskog between 1910 and 1920 (Chapman, 1916 and 1918; Enskog, 1917 and 1922; see also Cercignani, 1988 and 1990). The calculations for certain models are rather involved and may easily hide some peculiarities of this expansion. Therefore it seems appropriate to discuss a few interesting features before beginning with the formal derivations and to restrict the calculation to a simple collision model, namely the BGK approximation.

The Chapman-Enskog or multi-scale expansion has already been used to derive the Euler equation for the FHP lattice-gas cellular automata (compare Section 3.2) and will be applied later on to derive the Navier-Stokes and other macroscopic equations for lattice Boltzmann models (compare Section 5.2).

The transformation from the Boltzmann equation to the Navier-Stokes equation involves a *contraction of the description* of the temporal development of the system (Uhlenbeck and Ford, 1963). Whereas the distribution function f of the Boltzmann equation in general is *explicitly* depending on time, space and velocity, we will see that the distribution functions $f^{(n)}$ of the Chapman-Enskog expansion depend only *implicitly* on time via the local density, velocity and temperature, i.e. the $f^{(n)}$ are not the most general solutions of the Boltzmann equation. It can be shown that arbitrary initial distributions relax very fast (a few collision time scales which means of the order of 10^{-11} s in a gas in 3D under standard conditions) toward this special kind of distribution. The possibility of the contraction of the description has been considered as a *very fundamental insight* (Uhlenbeck and Ford, 1963).

The expansion parameter of Chapman-Enskog is the *Knudsen number* K_n, i.e. the ratio between the *mean free length* λ (the mean distance between two

succesive collisions) and the characteristic spatial scale of the system (for example, radius of an obstacle in a flow). When the Knudsen number is of the order of 1 or larger the gas in the system under consideration cannot be described as a fluid.

As a last point one should mention that the series resulting from the Chapman-Enskog procedure is probably not convergent but asymptotic[2]. This is suggested by the application to the dispersion of sound (Uhlenbeck and Ford, 1963). Higher order approximations of the Chapman-Enskog method lead to the Burnett and super-Burnett equations (Burnett, 1935, 1936) which have never been applied systematically. One of the problems with these equations is the question of appropriate boundary conditions (see, for example, Cercignani, 1988 and 1990, for further discussion).

4.2.1 The conservation laws

Conservation laws can be obtained by multiplying the Boltzmann equation with a collision invariant $\psi_k(\boldsymbol{v})$ ($\psi_0 = 1$, $\psi_\alpha = u_\alpha$ for $\alpha = 1, 2, 3$ and $\psi_4' = \frac{1}{2}m|\boldsymbol{v} - \boldsymbol{u}|^2$) and subsequent integration over d^3v. The integrals over the collision integral $Q(f, f)$ vanish by definition. Therefore

$$\int d^3v\, \psi_k(\partial_t + v_\alpha \partial_{x_\alpha}) f(\boldsymbol{x}, \boldsymbol{v}, t) = 0 \tag{4.2.1}$$

and thus (in 3D)

$$\partial_t \rho + \partial_{x_\alpha}(\rho u_\alpha) = 0 \tag{4.2.2}$$

$$\rho \partial_t u_\alpha + \rho u_\beta \partial_{x_\beta} u_\alpha = -\partial_{x_\alpha} \hat{P}_{\alpha\beta} \tag{4.2.3}$$

$$\rho \partial_t \theta + \rho u_\beta \partial_{x_\beta} \theta = -\frac{2}{3}\partial_{x_\alpha} q_\alpha - \frac{2}{3}\hat{P}_{\alpha\beta}\Lambda_{\alpha\beta} \tag{4.2.4}$$

with

$$n(\boldsymbol{x}, t) = \int d^3v\, f(\boldsymbol{x}, \boldsymbol{v}, t) \tag{4.2.5}$$

$$\rho(\boldsymbol{x}, t) = m\, n(\boldsymbol{x}, t) \quad (m = const) \tag{4.2.6}$$

$$\tag{4.2.7}$$

$$\rho u_\alpha(\boldsymbol{x}, t) = m \int d^3v\, v_\alpha\, f(\boldsymbol{x}, \boldsymbol{v}, t) \tag{4.2.8}$$

[2] Asymptotic series are discussed, for example, in Bender and Orszag (1978). Despite their missing convergence these series can be extremely useful. Bender and Orszag give a number of neat examples.

$$\theta(\boldsymbol{x},t) = k_B T(\boldsymbol{x},t) = \frac{m}{3n}\int d^3v\,(v_\alpha - u_\alpha)(v_\alpha - u_\alpha)\,f(\boldsymbol{x},\boldsymbol{v},t) \tag{4.2.9}$$

$$\Lambda_{\alpha\beta} = \frac{m}{2}(\partial_{x_\beta} u_\alpha + \partial_{x_\alpha} u_\beta) \tag{4.2.10}$$

$$\hat{P}_{\alpha\beta} = m\int d^3v\,(v_\alpha - u_\alpha)(v_\beta - u_\beta)\,f(\boldsymbol{x},\boldsymbol{v},t) \tag{4.2.11}$$

$$q_\alpha(\boldsymbol{x},t) = \frac{m^2}{2}\int d^3v\,(v_\alpha - u_\alpha)(v_\beta - u_\beta)(v_\beta - u_\beta)\,f(\boldsymbol{x},\boldsymbol{v},t) \tag{4.2.12}$$

Although the conservation equations are exact they are useless until one can solve the Boltzmann equation and apply the solution f to calculate (4.2.5) to (4.2.12). Please note that $\hat{P}_{\alpha\beta}$ is different from the momentum flux tensor introduced in Eq. (3.2.54) in that it does not contain the advection term.

4.2.2 The Euler equation

Inserting $f^{(0)} = f^{(M)}$ (the Maxwell distribution, compare Eq. 4.1.5) into Eqs. (4.2.5) to (4.2.12) leads to the following approximation of the conservation laws

$$\begin{aligned}
\partial_t \rho + \partial_{x_\alpha}(\rho u_\alpha) &= 0 \quad \text{(continuity equation)} \\
\rho\partial_t u_\alpha + \rho u_\beta \partial_{x_\beta} u_\alpha &= -\partial_{x_\alpha} p \quad \text{(Euler equation)} \\
\partial_t \theta + u_\beta \partial_{x_\beta}\theta &= -\frac{1}{c_v}\theta\,\partial_{x_\alpha} u_\alpha
\end{aligned}$$

where $p = nk_BT = n\theta$ is the pressure and $c_v = 3/2$ is the heat capacity at constant volume. The heat flux $\boldsymbol{q}$ vanishes in this approximation. The continuity equation is already in its final form. The dissipative terms in the equation of motion have to be derived from higher order approximation.

4.2.3 Chapman-Enskog expansion

The distribution function is expanded as follows

$$f = f^{(0)} + \epsilon f^{(1)} + \epsilon^2 f^{(2)} + ... \tag{4.2.13}$$

The symbol ϵ is often used in two different ways:

1. One speaks of an expansion as a power series in the small quantity ϵ, i.e. $|\epsilon| \ll 1$. In the case of Chapman-Enskog the Knudsen number K_n can be considered as the small expansion parameter.
2. The formal parameter ϵ in the expansions allows one to keep track of the *relative* orders of magnitude of the various terms. It will be considered only as a *label* and will be dropped out of the final results by setting $\epsilon = 1$.

As an example consider the expansion $f = f^{(0)} + \epsilon f^{(1)}$. In discussions one may consider $f^{(0)}$ and $f^{(1)}$ as quantities of the same order of magnitude and argue that the second term of the expansion is small because ϵ is a small quantity whereas in the formal calculations $f^{(1)}$ is small compared to $f^{(0)}$ and ϵ is only a label to keep track of the relative size of the various terms. The ϵ in this second sense can be set equal to one after finishing all transformations. According to the expansion (4.2.13) the conservation laws (Eqs. 4.2.2 - 4.2.4) can be formulated as follows

$$\begin{aligned}
\partial_t \rho + \partial_{x_\alpha}(\rho u_\alpha) &= 0 \\
\rho \partial_t u_\alpha + \rho u_\beta \partial_{x_\beta} u_\alpha &= -\sum_{n=0}^{\infty} \epsilon^n \partial_{x_\alpha} \hat{P}^{(n)}_{\alpha\beta} \\
\rho \partial_t \theta + \rho u_\beta \partial_{x_\beta} \theta &= -\frac{2}{3} \sum_{n=0}^{\infty} \epsilon^n (\partial_{x_\alpha} q^{(n)}_\alpha + \hat{P}^{(n)}_{\alpha\beta} \Lambda_{\alpha\beta})
\end{aligned}$$

where

$$\hat{P}^{(n)}_{\alpha\beta} := m \int d^3v \, f^{(n)} (v_\alpha - u_\alpha)(v_\beta - u_\beta) \tag{4.2.14}$$

and

$$q^{(n)}_\alpha := \frac{m^2}{2} \int d^3v \, f^{(n)} (v_\alpha - u_\alpha) |\boldsymbol{v} - \boldsymbol{u}|^2.$$

Because f depends on t only via ρ, $\boldsymbol{u}$ and T the chain rule

$$\partial_t f = \partial_\rho f \, \partial_t \rho + \partial_{u_\alpha} f \, \partial_t u_\alpha + \partial_\theta f \, \partial_t \theta$$

applies. Inserting (4.2.13) into the derivatives of f with respect to ρ, u_α and T yields

$$\begin{array}{rcccccccc}
\partial_\rho f &=& \partial_\rho f^{(0)} &+& \epsilon \partial_\rho f^{(1)} &+& \epsilon^2 \partial_\rho f^{(2)} &+& \ldots \\
\partial_{u_\alpha} f &=& \partial_{u_\alpha} f^{(0)} &+& \epsilon \partial_{u_\alpha} f^{(1)} &+& \epsilon^2 \partial_{u_\alpha} f^{(2)} &+& \ldots \\
\partial_\theta f &=& \partial_\theta f^{(0)} &+& \epsilon \partial_\theta f^{(1)} &+& \epsilon^2 \partial_\theta f^{(2)} &+& \ldots
\end{array}$$

The expansions of $\partial_t \rho$, $\partial_t u_\alpha$ and $\partial_t T$ have to be defined such that they are consistent with the conservation laws in each order of ϵ. The terms of the formal expansion[3]

$$\partial_t = \epsilon \partial_t^{(1)} + \epsilon^2 \partial_t^{(2)} + ... \tag{4.2.15}$$

will be derived from the conservation laws as follows:

$$\begin{aligned}
\partial_t^{(1)} \rho &:= -\partial_{x_\alpha}(\rho u_\alpha) && (4.2.16)\\
\partial_t^{(n+1)} \rho &:= 0 \quad (n > 0) \\
\partial_t^{(1)} u_\alpha &:= -u_\beta \partial_{x_\beta} u_\alpha - \frac{1}{\rho} \partial_{x_\beta} \hat{P}^{(0)}_{\alpha\beta} && (4.2.17)\\
\partial_t^{(n+1)} u_\alpha &:= -\frac{1}{\rho} \partial_{x_\beta} \hat{P}^{(n)}_{\alpha\beta} \quad (n > 0) \\
\partial_t^{(1)} \theta &:= -u_\beta \partial_{x_\beta} \theta - \frac{2}{3\rho} \left(\partial_{x_\alpha} q^{(0)}_\alpha + \hat{P}^{(0)}_{\alpha\beta} \Lambda_{\alpha\beta} \right) \\
\partial_t^{(n+1)} \theta &:= -\frac{2}{3\rho} \left(\partial_{x_\alpha} q^{(n)}_\alpha + \hat{P}^{(n)}_{\alpha\beta} \Lambda_{\alpha\beta} \right) \quad (n > 0)
\end{aligned}$$

Application of these definitions leads to an expansion of $\partial_t f$ into a power series in ϵ:

$$\begin{aligned}
\partial_t f &= \left(\epsilon \partial_t^{(1)} + \epsilon^2 \partial_t^{(2)} + ... \right) \left(f^{(0)} + \epsilon f^{(1)} + \epsilon^2 f^{(2)} + ... \right) \\
&= \epsilon \partial_t^{(1)} f^{(0)} + \epsilon^2 (\partial_t^{(1)} f^{(1)} + \partial_t^{(2)} f^{(0)}) + \epsilon^3 ...
\end{aligned}$$

Inserting the expansion of the distribution function f into the collision integral $Q(f, f)$ of the Boltzmann equation with BGK approximation[4] yields

$$\begin{aligned}
Q(f, f) &= -\omega \left(f - f^{(0)} \right) \\
&= -\omega \left(\epsilon f^{(1)} + \epsilon^2 f^{(2)} + ... \right) \\
&=: J^{(0)} + \epsilon J^{(1)} + \epsilon^2 J^{(2)} + ... && (4.2.18)
\end{aligned}$$

[3] The reason for starting the ∂_t expansion by a term linear in ϵ will become apparent from the discussion later on. The expansions of f or ∂_t alone can be multiplied by arbitrary powers of ϵ because the powers of ϵ only label the *relative* size of the different terms in each expansion. When expansions of different quantities are combined, however, the powers of ϵ have to be related such that the terms of leading order yield a meaningful balance.

[4] The BGK approximation will be applied here in order to simplify the calculations.

where

$$J^{(0)}\left(f^{(0)}\right) = 0 \tag{4.2.19}$$

$$J^{(1)}\left(f^{(0)}, f^{(1)}\right) = J^{(1)}\left(f^{(1)}\right) = -\omega f^{(1)} \tag{4.2.20}$$

$$J^{(2)}\left(f^{(0)}, f^{(1)}, f^{(2)}\right) = J^{(2)}\left(f^{(2)}\right) = -\omega f^{(2)} \tag{4.2.21}$$

$$\ldots$$

where the collision frequency ω is a constant. In general, i.e. no BGK approximation of the collision integral, the $J^{(n)}$ depend on all $f^{(k)}$ with $k \leq n$ (as indicated on the left hand sides of Eqs. 4.2.20 and 4.2.21) whereas for the BGK approximation $J^{(n)}$ depends only on $f^{(n)}$. This simplification is due to the fact that the collision integral in the BGK approximation is linear in f. The spatial derivative $\partial_{\boldsymbol{x}}$ on the left hand side of the Boltzmann equation is of the same order as the leading term in the time derivative, i.e.

$$\partial_{x_\alpha} = \epsilon \partial^{(1)}_{x_\alpha}. \tag{4.2.22}$$

This looks like the first term of an expansion. In space, however, only one macroscopic scale will be considered because different macroscopic processes like advection and diffusion can be distinguished by their time scales but act on similar spatial scales.

Equating terms of same order in ϵ of the Boltzmann equation leads to the following set of equations:

$$J^{(0)}\left(f^{(0)}\right) = 0 \tag{4.2.23}$$

$$\partial^{(1)}_t f^{(0)} + v_\alpha \partial^{(1)}_{x_\alpha} f^{(0)} = J^{(1)}\left(f^{(0)}, f^{(1)}\right) = -\omega f^{(1)} \tag{4.2.24}$$

$$\partial^{(1)}_t f^{(1)} + \partial^{(2)}_t f^{(0)} + v_\alpha \partial^{(1)}_{x_\alpha} f^{(1)} = J^{(2)}\left(f^{(0)}, f^{(1)}, f^{(2)}\right) = -\omega f^{(2)}$$

$$\ldots$$

Eq. (4.2.23) is fulfilled because J vanishes for Maxwell distributions. $f^{(1)}$ can readily be calculated from Eq. (4.2.24)

$$\boxed{f^{(1)} = -\frac{1}{\omega}\left(\partial^{(1)}_t f^{(0)} + v_\alpha \partial^{(1)}_{x_\alpha} f^{(0)}\right).} \tag{4.2.25}$$

This equation states that the lowest order deviations $f^{(1)}$ from a local Maxwell distribution $f^{(0)}$ are proportional to the gradient in space and time of $f^{(0)}$. The calculation of $f^{(1)}$ is much more involved when the collision integral is not approximated (see, for example, Huang, 1963).

The next step is the calculation of $\hat{P}^{(1)}_{\alpha\beta}$ according to Eq. (4.2.14)

$$\begin{aligned}\hat{P}^{(1)}_{\alpha\beta} &= m\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)f^{(1)}\\ &= -\frac{m}{\omega}\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)(\partial^{(1)}_t f^{(0)}+v_\gamma\partial^{(1)}_{x_\gamma}f^{(0)}).\end{aligned}$$

Insertion of (4.2.16) and (4.2.17) leads to (from now on the superscript $^{(1)}$ will be dropped for the sake of simplicity)

$$\begin{aligned}\partial_t f^{(0)}(\rho,\boldsymbol{u}) &= \frac{\partial f^{(0)}}{\partial\rho}\frac{\partial\rho}{\partial t}+\frac{\partial f^{(0)}}{\partial u_\gamma}\frac{\partial u_\gamma}{\partial t}\\ &= -\frac{f^{(0)}}{m}\frac{\partial(\rho u_\gamma)}{\partial x_\gamma}+\frac{m}{k_BT}(v_\gamma-u_\gamma)f^{(0)}\left(u_\delta\frac{\partial u_\gamma}{\partial x_\delta}+\frac{1}{\rho}\frac{\partial\hat{P}^{(0)}_{\gamma\delta}}{\partial x_\delta}\right)\\ &= -\frac{\rho f^{(0)}}{m}\frac{\partial u_\gamma}{\partial x_\gamma}-\frac{f^{(0)}}{m}u_\gamma\frac{\partial\rho}{\partial x_\gamma}+\frac{m}{k_BT}(v_\gamma-u_\gamma)f^{(0)}u_\delta\frac{\partial u_\gamma}{\partial x_\delta}\\ &\quad+\frac{m}{k_BT}(v_\gamma-u_\gamma)f^{(0)}\frac{1}{\rho}\delta_{\gamma\delta}\frac{\partial p}{\partial x_\delta}\end{aligned}$$

and

$$\begin{aligned}v_\gamma\partial_{x_\gamma}f^{(0)} &= v_\gamma\frac{\partial f^{(0)}}{\partial\rho}\frac{\partial\rho}{\partial x_\gamma}+v_\gamma\frac{\partial f^{(0)}}{\partial u_\delta}\frac{\partial u_\delta}{\partial x_\gamma}\\ &= v_\gamma\frac{f^{(0)}}{m}\frac{\partial\rho}{\partial x_\gamma}+v_\gamma\frac{m}{k_BT}(v_\delta-u_\delta)f^{(0)}\frac{\partial u_\delta}{\partial x_\gamma}.\end{aligned}$$

The various integrals are readily evaluated

$$-\frac{1}{m}\frac{\partial u_\gamma}{\partial x_\gamma}\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)f^{(0)}=-\delta_{\alpha\beta}\,n\frac{k_BT}{m}\frac{\partial u_\gamma}{\partial x_\gamma}$$

$$\frac{1}{m}\frac{\partial\rho}{\partial x_\gamma}\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)(v_\gamma-u_\gamma)f^{(0)}=0$$

$$\frac{m}{k_BT}f^{(0)}\left(u_\delta\frac{\partial u_\gamma}{\partial x_\delta}+\frac{1}{\rho}\delta_{\gamma\delta}\frac{\partial p}{\partial x_\delta}\right)\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)(v_\gamma-u_\gamma)f^{(0)}=0$$

$$\begin{aligned}&\frac{m}{k_BT}\frac{\partial u_\delta}{\partial x_\gamma}\int d^3v(v_\alpha-u_\alpha)(v_\beta-u_\beta)v_\gamma(v_\delta-u_\delta)f^{(0)}\\ =\;&(\delta_{\alpha\beta}\delta_{\gamma\delta}+\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})n\frac{k_BT}{m}\frac{\partial u_\delta}{\partial x_\gamma}\end{aligned}$$

and thus

$$\begin{aligned}\hat{P}^{(1)}_{\alpha\beta} &= -n\frac{k_BT}{\omega}\left[(\delta_{\alpha\beta}\delta_{\gamma\delta}+\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})\frac{\partial u_\delta}{\partial x_\gamma}-\delta_{\alpha\beta}\frac{\partial u_\gamma}{\partial x_\gamma}\right]\\ &= n\frac{k_BT}{\omega}\begin{pmatrix} 2\dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y}+\dfrac{\partial v}{\partial x} & \dfrac{\partial u}{\partial z}+\dfrac{\partial w}{\partial x}\\ \dfrac{\partial u}{\partial y}+\dfrac{\partial v}{\partial x} & 2\dfrac{\partial v}{\partial y} & \dfrac{\partial v}{\partial z}+\dfrac{\partial w}{\partial y}\\ \dfrac{\partial u}{\partial z}+\dfrac{\partial w}{\partial x} & \dfrac{\partial v}{\partial z}+\dfrac{\partial w}{\partial y} & 2\dfrac{\partial w}{\partial z}\end{pmatrix}.\end{aligned}$$

Neglecting density and temperature variations the divergence of $\hat{P}^{(1)}_{\alpha\beta}$ reads

$$\begin{aligned}\frac{\partial \hat{P}^{(1)}_{\alpha\beta}}{\partial x_\alpha} &= \mu\left[2\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}+\frac{\partial^2 v}{\partial x\partial y}+\frac{\partial^2 w}{\partial x\partial z}\right]\boldsymbol{e}_x+...\\ &= \mu\left[\frac{\partial}{\partial x_\beta}\left(\frac{\partial u_\alpha}{\partial x_\beta}\right)+\frac{\partial}{\partial x_\alpha}\left(\frac{\partial u_\beta}{\partial x_\beta}\right)\right]\\ &= \mu\left[\nabla^2\boldsymbol{u}+\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{u})\right]\end{aligned}$$

where

$$\mu = n\frac{k_BT}{\omega} \tag{4.2.26}$$

is the dynamic shear viscosity. Thus one obtains the Navier-Stokes equation

$$\boxed{\partial_t u_\alpha + u_\beta\partial_{x_\beta}u_\alpha = -\partial_{x_\alpha}P+\nu\partial_{x_\beta}\partial_{x_\beta}u_\alpha+\xi\partial_{x_\alpha}\partial_{x_\beta}u_\beta} \tag{4.2.27}$$

where the kinematic shear (ν) and bulk (ξ) viscosities are equal and given by

$$\nu = \frac{k_BT}{\omega m} = \xi. \tag{4.2.28}$$

4.3 The maximum entropy principle

In 1948 Shannon [418, 419] proposed a theory which allows us to quantify 'information'. The statistical measure of the lack of information is called the information theoretical or Shannon entropy. Equilibrium distributions can be derived from the maximum entropy principle.

The following presentation closely follows Stumpf and Rieckers (1976).
First consider a *discrete set* $Z := \{z_1 ... z_N\}$ with N elements. A *message*[5] is defined as a selection of one or several elements of Z. The informational measure of the message is defined by that knowledge which is necessary to denote a certain element or a selection of elements. What is the elementary unit of this measure? If the set Z encompasses only one element the selection of this element does not augments our knowledge. There is no real message until the number of elements in Z is at least two. Obviously the decision between two alternatives is the smallest unit of information one can think of: it is called a *bit* which is the short form of *'binary digit'*. The larger the number of elements in Z, the more information is connected with the selection of a certain element of Z. The measure of the information gained can be traced back to a sequence of alternative decisions. The number of elements N can be written down in binary form. The number of binary digits is a measure of information. Or the elements z_j can be aranged in the form of a binary tree (compare Fig. 4.3.1) where the number of branching points from the root to one of the end points equals the number of bits. These procedures work for sets with $N = 2^n$ elements and yield the measure of information $I(N) = n = \log_2 N$ for the selection of a single element. This definition is generalized to sets with arbitrary number of elements by

$$I(N) = \log_2 N,$$

i.e. $I(N)$ is not necessary an integer anymore.

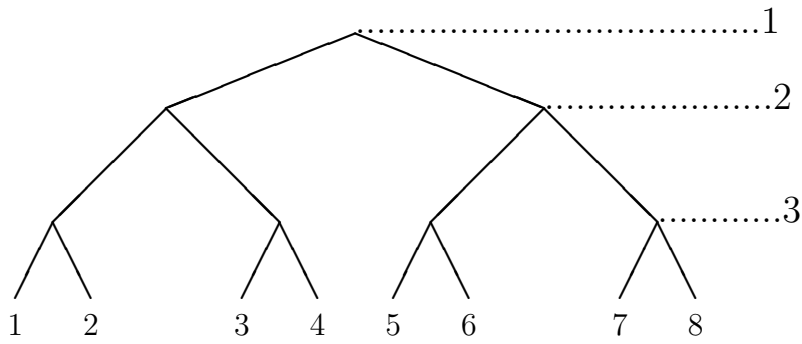
Further the measure of information is additive with respect to the choice of several (α) elements out of a set with N elements

$$I(N, \alpha) = \alpha \cdot I(N)$$

and the choice of two elements out of a direct product of two sets Z_1 and Z_2

[5] The notation has its roots in the theory of communication. One of the basic problems in this context is the reliable transmission of messages from a source via a channel to a receiver. Often the messages to be transmitted have meaning like, for example, the news you hear on the radio. This, however, is not always the case. In transmitting music, the meaning is much more subtle then in the case of a verbal message. In any case, meaning is quite irrelevant to the problem of transmitting the information.

Fig. 4.3.1. *The information for the selection of a certain element out of a set of* $N = 2^n$ *elements is defined as the number of alternative decisions necessary when going from the root to a certain end point. The selection of a certain elements out of* 8 *elements requires three binary decisions.*



$$I(N_{Z_1 \otimes Z_2}) = I(N_{Z_1}) + I(N_{Z_2}).$$

Now consider *probability distributions* instead of sets. Let us start with discrete probability distributions P with a finite number of entries:

$$P := \{P_1 ... P_N\}\,, \quad \sum_{k=1}^{N} P_k = 1$$

corresponding to the set of events

$$\Omega := \{x_1 ... x_N\}\,.$$

The task is to find a measure of information $I(P)$ for a probability distribution as a whole. Let's first consider two special cases:

– The *sharp distribution*: $P_i = \delta_{il}$, i.e. each measurement will yield the result x_l. The sharp distribution contains a lot of information: if you know this probability distribution you can be sure of the output of your next measurement. Because one event out of N possible events is selected the measure

$$I(P) = \log_2 N$$

suggests itself.

– The *normal distribution*: $P_i = 1/N$, i.e. every possible event has the same probability. The normal distribution can be understood as a consequence of the *Laplacian principle of the insufficient reason*: "If there is no reason to single out a certain event for given information concerning an experimental situation, a normal distribution is to be assumed." (Stumpf and Rieckers, 1976, p.14). Obviously the normal distribution contains the minimal amount of information, thus

$$I(P) = 0.$$

The measure of information $I(P)$ for a general distribution $I(P) = I(P_1...P_N)$ is based on four postulates:

1. $I(P)$ is an *universal function*.
2. $I(P)$ is additive concerning the (still to be determined) statistical information for single elements $i(P_k)$ as well as for the composition of direct products:
$$I(P_1...P_N) = \sum_{k=1}^{N} i(P_k)$$
$$I(P_{\Omega_1 \otimes \Omega_2}) = I(P_{\Omega_1}) + I(P_{\Omega_2}).$$
3. $I(P) = 0$ for the normal distribution and $I(P) = \log_2 N$ for the sharp distribution.
4. The statistical information of a single element $i(P)$ is defined on $0 \leq P \leq 1$ and is continuous.

Theorem 4.3.1. *The statistical measure of information $I(P)$ over the set of events Ω with N elements, which fulfills the above given postulates 1 to 4, is uniquely given by*

$$I(P) = I(P_1...P_N) = \sum_{i=1}^{N} P_i \log_2 P_i + \log_2 N. \tag{4.3.1}$$

The proof can be found, for example, in Stumpf and Rieckers (1976).

Lemma 4.3.1. *The maximum value I_{max} of $I(P)$ is given by the sharp distribution:*

$$I_{max}(P) = \log_2 N.$$

Exercise 4.3.1. (*)
Prove Lemma 4.3.1.

The *information theoretical entropy* or *Shannon entropy* S is defined as follows:

$$S(P_1...P_N) := I_{max} - I(P_1...P_N) = -\sum_{i=1}^{N} P_i \log_2 P_i. \tag{4.3.2}$$

It is a measure for the *lack of information*: S vanishes for the sharp distribution and becomes maximal for the normal distribution.
The generalization of $I(P)$ for continuous sets of events is given by

$$S[f] := -k \int f(x) \ln f(x)\, dx, \tag{4.3.3}$$

i.e. the function $S = S(P)$ is replaced by a functional $S = S[f]$ over the probability density f. The transition from the case of discrete distributions to probability densities f is not as simple as it looks. For example, there is no maximal measure of information and because f can be a generalized function the integral 4.3.3 could be meaningless (see Stumpf and Rieckers, 1976, for a detailed discussion).

The most important theorem of this section reads:

Theorem 4.3.2. *(Maximum entropy principle) If the probability density $f(x)$ with the normalization*

$$\int f(x)\, dx = 1$$

obeys the following m linear independent constraints

$$\int R_i(x) f(x)\, dx = r_i \quad 1 \le i \le m$$

then the probability density which maximizes the lack of information while respecting the $m+1$ constraints is uniquely given by

$$f(x) = \exp\left[-\lambda_0 - \sum_{i=1}^{m} \lambda_i R_i(x)\right] \tag{4.3.4}$$

where the Lagrange multipliers λ_0, λ_1, ..., λ_m are unique functions of the values $r_1, ..., r_m$.

Proof. The proof essentially consists of two parts. Here only the derivation of the distribution (4.3.4) shall be discussed in detail. The second part, namely the proof that the Lagrange multipliers are uniquely determined by the values $r_1, ..., r_m$, can be found in Stumpf and Rieckers (1976, p. 20).

The extremum of a functional under given constraints is sought after. An extended functional $\hat{S}[f]$ is defined by coupling the constraints via Lagrange multipliers $\eta_0, ...\eta_m$ to the Shannon entropy $S[f]$:

$$\begin{aligned}\hat{S}[f] &:= S[f] - k\,(\lambda_0 - 1)\; Tr[f] - k\sum_{i=1}^{m}\lambda_i\, Tr[(R_i - r_i)f] \\ &= -k\, Tr\left[f\left(\ln f + \lambda_0 - 1 + \sum_{i=1}^{m}\lambda_i(R_i - r_i)\right)\right].\end{aligned}$$

For reasons which become obvious in a moment the η_j have been written as follows:

$$\eta_0 = k(\lambda_0 - 1)$$
$$\eta_i = k\lambda_i \quad 1 \le i \le m.$$

The *trace* of f is defined as

$$Tr[f] := \int f(x)\, dx. \tag{4.3.5}$$

From this definition it immediately follows

$$Tr[c \cdot f] = \int c \cdot f(x)\, dx = c \int f(x)\, dx = c\, Tr[f]$$

where c is a constant.

The vanishing of the *functional derivative* of $\hat{S}[f]$ with respect to f is a necessary condition for an extremum of $S[f]$:

$$\frac{\delta \hat{S}[f]}{\delta f} = 0$$

Functional derivatives are calculated analogously to the rules for ordinary derivatives (see, for example, Großmann, 1988):

$$\frac{\delta \hat{S}[f]}{\delta f} = -k\left\{\ln f + \lambda_0 + \sum_{i=1}^{m}\lambda_i R_i\right\}$$

and therefore

$$\ln f = -\lambda_0 - \sum_{i=1}^{m}\lambda_i R_i$$

respectively

$$f = \exp\left[-\lambda_0 - \sum_{i=1}^{m}\lambda_i R_i\right].$$

q.e.d.

The maximum entropy principle will be applied later on to calculate equilibrium distributions for lattice Boltzmann models.
Further reading: The proceedings edited by Levine and Tribus (1979) and especially the paper by Jaynes (1979).

Exercise 4.3.2. (**)
Find s_i $(i = 0, 1, ..., l)$ such that

$$\sum_i \boldsymbol{c}_i s_i(\boldsymbol{x}, t) = \boldsymbol{S}(\boldsymbol{x}, t)$$

under the constraint

$$\sum_i s_i(\boldsymbol{x}, t) = 0$$

by minimizing

$$V = \sum s_i^2.$$

The lattice velocities $\boldsymbol{c}_i$ satisfy

$$\sum_i \boldsymbol{c}_i = 0, \quad \text{and} \quad \sum_i \boldsymbol{c}_i^2 = n.$$

Exercise 4.3.3. (**)
The *Renyi entropy* (Renyi, 1970) of order α is definiered as follows:

$$S_\alpha := -\frac{1}{\alpha - 1} \ln \sum_{i=1}^{N} p_i^\alpha, \quad \alpha \in \mathcal{R}, \quad \alpha \neq 1.$$

Calculate

$$\lim_{\alpha \to 1} S_\alpha.$$