Reconstruction of the primordial Universe by a Monge-Ampère-Kantorovich optimisation scheme

Uriel Frisch[1], Sabino Matarrese[2], Roya Mohayaee[1] and Andrei Sobolevskiĭ[1,3]
[1]*Département Cassini, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice, Cedex 4, France*
[2] *Dipartimento di Fisica "Galileo Galilei" and INFN, Sezione di Padova, via Marzolo 8, I-35131 Padova, Italy*
[3] *Department of Physics, M V Lomonossov University, 119899-Moscow, Russia*

A new method for the reconstruction of primordial density fluctuation field is presented [U. Frisch, S. Matarrese, R. Mohayaee and A. Sobolevskii, 2002, Nature 417, 260]. Various previous approaches to this problem rendered *non-unique* solutions. We have shown that, under a suitable formulation, reconstruction is a well-posed problem. It is demonstrated that initial positions of dark matter fluid elements, under the hypothesis that their displacement is the gradient of a convex potential, can be reconstructed uniquely. We have shown that the cosmological reconstruction problem can be formulated as an assignment problem in optimisation theory. When tested against numerical simulations, our scheme yields excellent reconstruction on scales larger than a few megaparces.

Based on a talk presented by Roya Mohayaee at Marseille colloque de PNC, Luminy, September 2002

I. INTRODUCTION

The present distribution of galaxies brought to us by redshift surveys indicates that the Universe on large scales exhibits a high degree of clumpiness with coherent structures such as voids, great walls, filaments and clusters. The cosmic microwave background (CMB) explorers, however, indicate that the Universe was highly homogeneous billions of years ago. When studying these data, among the questions that are of concern in cosmology are the initial conditions of the Universe and the dynamics under which it grew into the present Universe. CMB explorers provide us with valuable knowledge into the initial conditions of the Universe, but the present distribution of the galaxies opens a second, complementary window into the early Universe.

Unravelling the properties of the early Universe from the present data is an instance of the general class of *inverse problems* in physics. The orthodox method is to tackle this problem in an empirical way by taking a *forward approach.* In the forward approach, a cosmological model is proposed for the initial power spectrum of dark matter. Next, a particle presentation of the initial density field is made which provides the initial data for a N-body simulation which is run using Newtonian dynamics and is stopped at the present time. Subsequently, a *statistical* comparison between the outcome of the simulation and the observational data can be made assuming that a suitable *bias* relation exists between the distribution of galaxies and that of dark matter. If the statistical test is satisfactory then the implication is that the initial condition was viable, otherwise one changes the cosmological parameters and goes through the whole process again. This is repeated until one obtains a satisfactory statistical test, affirming a good choice for the initial condition.

However, this inverse problem does not have to be necessarily dealt with in a forward manner as explained above and can be tackled differently. Can one fit the present distribution of the galaxies *exactly* rather than statistically and run it back in time to make the *reconstruction* of the primordial density fluctuation field? Since Newtonian gravity is time-reversible, one would have been able to integrate the equations of motions back in time and solve the reconstruction problem trivially, if in addition to their positions, the present velocities of the galaxies were also known. As a matter of fact, however, the peculiar velocities of only a few thousands of galaxies are known out of the hundreds of thousands whose redshifts have been measured. Indeed, one goal of reconstruction is to evaluate the peculiar velocities of the galaxies and in this manner put direct constraints on cosmological parameters.

Without a second boundary condition, reconstruction would thus be an infeasible task. Newton's equation of motion requires two boundary conditions, whereas, for reconstruction so far we only have mentioned the present positions of the galaxies. The second condition is the homogeneity of the initial density field: as we go back in time the peculiar velocities of the galaxies vanish. Thus, contrary to the forward approach where one solves an *initial-value problem*, in the reconstruction approach one is dealing with a *two-point mixed boundary value problem.* In the former, one starts with the initial positions and velocities of the particles and solves Newton's equations arriving at a *unique* final position and velocity for a given particle. In the latter one does not always have *uniqueness.* This has been one of the shortcomings of reconstruction, which was consequently taken to be an ill-posed problem.

II. VARIATIONAL APPROACH TO RECONSTRUCTION

The history of reconstruction goes back to the work of Peebles (Peebles 1989) on tracing the orbits of the members of the local group. In his approach, reconstruction was solved as a variational problem. Instead of solving the equations of motion, one searches for the stationary points of the corresponding Euler-Lagrange action. The action in the comoving coordinates is (Peebles 1980)

$$S = \int \sum_i \left[\frac{m_i a^2 \dot{\mathbf{x}}_i^2}{2} - \sum_{j \neq i} \frac{G m_i m_j}{a|\mathbf{x}_i - \mathbf{x}_j|} + \frac{2}{3}\pi G \rho_{\rm b} a^2 m_i \mathbf{x}_i^2 \right] \tag{1}$$

where the path of the ith particle with mass m_i is $\mathbf{x}_i(t)$, $\rho_{\rm b}$ is the mean mass density, and the present value of the expansion parameter $a(t)$ is $a_0 = a(t_0) = 1$. The equation of motion is obtained by requiring

$$\delta S = \int dt \left[\frac{\partial \mathcal{L}}{\partial \mathbf{x}_i} \cdot \delta \mathbf{x}_i - \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}_i} \cdot \delta \dot{\mathbf{x}}_i \right] = 0 \tag{2}$$

which leads to

$$\int dt \delta \mathbf{x}_i \left[\sum_{j \neq i} \frac{G m_i m_j (\mathbf{x}_i - \mathbf{x}_j)}{a|\mathbf{x}_i - \mathbf{x}_j|^3} + \frac{4}{3}\pi G \rho a^2 m_i \mathbf{x}_i - m_i \frac{\partial}{\partial t}(a^2 \dot{\mathbf{x}}_i^2) \right] - m_i \left[a^2 \dot{\mathbf{x}}_i \cdot \delta \mathbf{x}_i \right]_0^{t_0} = 0, \tag{3}$$

where t_0 denotes the present time. The mixed boundary conditions

$$\begin{aligned} \delta \mathbf{x}_i = 0 \qquad & \text{at} \qquad t = t_0 \\ a^2 \dot{\mathbf{x}}_i = 0 \qquad & \text{at} \qquad t = 0 \end{aligned} \tag{4}$$

would then eliminate the boundary terms in (3).

The components $\alpha = 1, 2, 3$ of the orbit of the ith particle are modelled as

$$x_i^\alpha = x_i^\alpha(t_0) + \sum_n C_{i,n}^\alpha f_n(t) \tag{5}$$

The functions f_n are normally convenient functions of the scale factor a and should satisfy the boundary conditions (4). Initially, trial functions such as

$$f_n = a^n(1 - a) \tag{6}$$

or

$$f_n = \sin\left(\frac{n\pi a}{2}\right) \tag{7}$$

and $f_0 = \cos(\pi a/2)$ were used. The coefficients $C_{i,n}$ are then found by substituting expression (5) in the action (1) and finding the stationary points. That is for physical trajectories

$$\frac{\partial S}{\partial C_{i,n}} = 0 \tag{8}$$

leading to

$$m_i \int dt f_n(t) \left[-\frac{d}{dt} a^2 \frac{dx_i^\alpha}{dt} + \frac{4}{3}\pi G \rho_{\rm b} a^2 x_i^\alpha(t) \right] = -\frac{G}{a} \int_0^{t_0} \sum_j m_j \frac{x_j^\alpha - x_i^\alpha}{|\mathbf{x}_i - \mathbf{x}_j|} \tag{9}$$

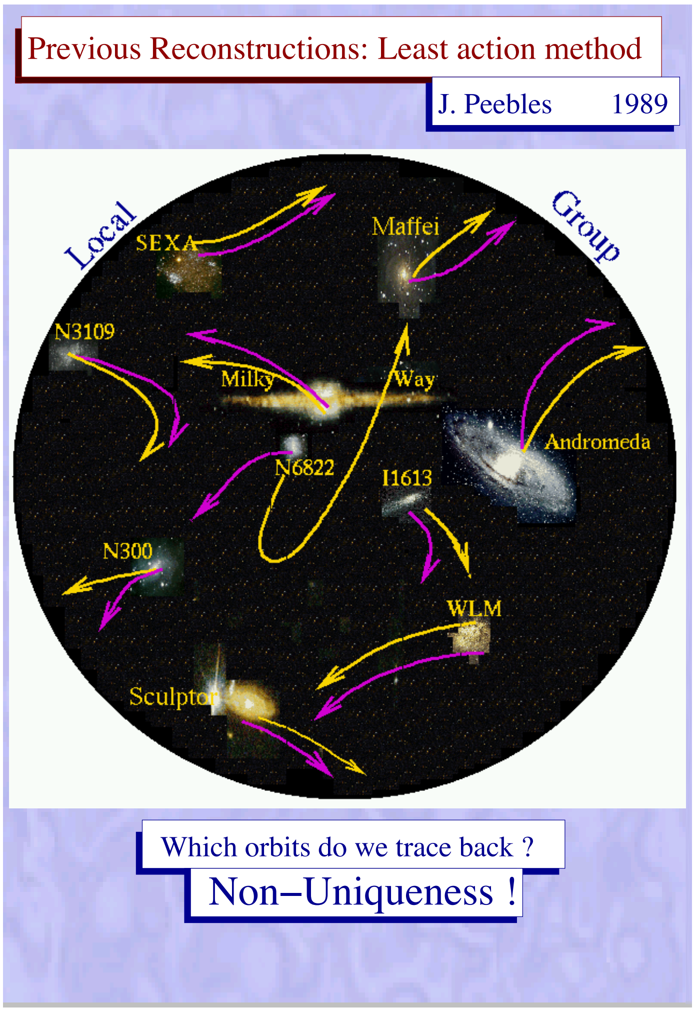

FIG. 1. A schematic demonstration of Peebles' reconstruction of the trajectories of the members of the local group using a variational approach based on the minimisation of Euler-Lagrange action. In most cases there is more than one allowed trajectory due to orbit crossing (closely related to the multi-streaming of the underlying dark matter fluid). The pink (darker) orbits correspond to taking the minimum of the action whereas the yellow (brighter) orbits were obtained by taking the saddle-point solution. Of particular interest is the orbit of N6822 which in the former solution is on it first approach towards up and in the second solution is in its passing orbit. A better agreement between the evaluated and observed velocities was shown to correspond to the saddle point solution.

In his first work, Peebles (1989) considered only the minimum of the action while reconstructing the trajectories of the galaxies back in time. In a low-density Universe assuming a linear bias, the predicted velocities agreed with the observed ones for most galaxies of the local group but failed with a large discrepancy for remaining members. Later on, it was found that if the trajectories corresponding to the saddle-point of the action were taken instead of the minimum, much better an agreement between predicted and observed velocities would be obtained, for almost all the galaxies which were studied (see Fig. 1). Thus, by adjusting the orbits until the predicted and observed velocities agreed, reasonable bounds on cosmological parameters were found (Peebles 1989).

Although rather successful, reconstruction as such could not be applied to large galaxy redshift surveys containing hundreds of thousands of galaxies for the majority of which the peculiar velocities are unknown. Then, it is not possible to use the velocities, to choose the right orbit from the many which are all physically possible. In order to resolve the problem of multiple solutions (the existence of many physically possible orbits) one normally had to do significant smoothing and then try the computationally-costly reconstruction using Peebles variational approach (Shaya et al. 1995, Branchini et al. 2001). However, one was still not guaranteed to have chosen the right orbit.

We have shown that under a certain formulation reconstruction is a well-posed problem which is guaranteed to have a unique solution. The multiple solution can be caused by various factors. For example, the discretisation in the numerical integrations (9) can produce spurious valleys in the landscape of the Euler-Lagrange action. However, even overcoming all these numerical artifacts one is still not guaranteed uniqueness. There is a genuine physical reason for the lack of uniqueness which is what is often referred to in cosmology as *multi-streaming*. Cold dark matter (CDM) is a collisionless fluid with no velocity dispersion. An important feature that arises in the course of the evolution of a self-gravitating CDM universe is the formation of caustics: manifolds on which the velocity field is non-unique and the density is divergent. Regions bounded by these manifolds are often referred to as multi-stream regions. At the multi-stream point where velocity is multiple-valued, a particle can have many different physically viable trajectories each of which would correspond to a different stationary point of the Euler-Lagrange action which is no-longer *convex*.

III. MONGE-AMPÈRE-KANTOROVICH RECONSTRUCTION

Thus, reconstruction is a well-posed problem for as long as we avoid multi-stream regions. The mathematical formulation of this problem is as follows (Frisch et al. 2002). Unlike most of the previous works on reconstruction where one studies the Euler-Lagrange action, we start from a constraint equation, namely the mass conservation,

$$\rho(\mathbf{x})d\mathbf{x} = \rho(\mathbf{q})d\mathbf{q} \ , \tag{10}$$

where $\rho_0(\mathbf{q})$ is the density at the initial position, $\mathbf{q}$, and $\rho(\mathbf{x})$ is the density at the present position, $\mathbf{x}$, of the fluid element. The above mass conservation equation can be rearranged in the following form

$$\det \left[\frac{\partial q_i}{\partial x_j}\right] = \frac{\rho(\mathbf{x})}{\rho_0(\mathbf{q})} \tag{11}$$

where det stands for determinant. The right-hand-side of the above expression is basically given by our boundary conditions: the final positions of the particles are known and the initial distribution is homogeneous, $\rho_0(\mathbf{q} = \text{const.}$ To solve the equation, we make the following two hypotheses: the Lagrangian map ($\mathbf{q} \to \mathbf{x}$), is the gradient of a convex potential Φ. That is

$$\mathbf{x}(\mathbf{q}, t) = \nabla_q \Phi(\mathbf{q}, t). \tag{12}$$

The convexity guarantees that a single Lagrangian position corresponds to a single Eulerian position, *i.e.*, there has been no multi-streaming [1]. These assumptions imply that the inverse map $\mathbf{x} \to \mathbf{q}$ has also a potential representation

$$\mathbf{q} = \nabla_{\mathbf{x}} \Theta(\mathbf{x}, t) \tag{13}$$

where the potential $\Theta(\mathbf{x})$ is also a convex function and is related to $\Phi(\mathbf{x})$ by the Legendre-Fenchel transform

$$\Theta(\mathbf{x}) = \max_{\mathbf{q}} \left[\mathbf{q} \cdot \mathbf{x} - \Phi(\mathbf{q})\right] \quad ; \quad \Phi(\mathbf{q}) = \max_{\mathbf{x}} \left[\mathbf{x} \cdot \mathbf{q} - \Theta(\mathbf{x})\right] \tag{14}$$

The inverse map is now substituted in (11) yielding

$$\det \left[\frac{\partial^2 \Theta(\mathbf{x}, t)}{\partial x_i \partial x_j}\right] = \frac{\rho(\mathbf{x})}{\rho_0(\mathbf{q})} \tag{15}$$

which is the well-known Monge-Ampère equation. The solution to this 200 years old problem has recently been discovered (Brenier 1987) when it was realised that the map generated by the solution to the Monge-Ampère equation is the unique solution to an optimisation problem. This is the Monge-Kantorovich mass transportation problem, in which one seeks the map $\mathbf{x} \to \mathbf{q}$ which minimises the quadratic *cost* function

$$I = \int_{\mathbf{q}} \rho_0 |\mathbf{x} - \mathbf{q}|^2 d^3q = \int_{\mathbf{x}} \rho_0 |\mathbf{x} - \mathbf{q}|^2 d^3x \ . \tag{16}$$

A sketch of the proof is as follows. A small variation in the cost function yields

---

[1] The gradient condition has been made in previous works (Bertschinger and Dekel 1989) on the reconstruction of the peculiar velocities of the galaxies using linear Lagrangian theory.

$$\delta I = \int_{\mathbf{x}} \left[2\rho(\mathbf{x})(\mathbf{x} - \mathbf{q}) \cdot \delta \mathbf{x} \right] d^3 x \tag{17}$$

which must be supplemented by the condition

$$\nabla_{\mathbf{x}} \cdot (\rho(\mathbf{x}) \delta \mathbf{x}) = 0 \tag{18}$$

which expresses the cosntraint that the Eulerian density remains unchanged. The vanishing of δI should then hold for all $\mathbf{x} - \mathbf{q}$ which are orthogonal (in L^2) to functions of zero divergence. These are clearly gradients. Hence $\mathbf{x} - \mathbf{q}(\mathbf{x})$ and thus $\mathbf{q}(\mathbf{x})$ is a gradient of a function of $\mathbf{x}$.

Discretising the cost (16) yields

$$I = \mathrm{Min} \sum_{i=1}^{N} \left(\mathbf{q}_{j(i)} - \mathbf{x}_i \right)^2 \tag{19}$$

where the minimum is taken over all permutations $j(i)$. The formulation presented in (19) is known as the *the assignment problem*: given N initial and N final entries one has to find the permutation which minimises the quadratic cost function.

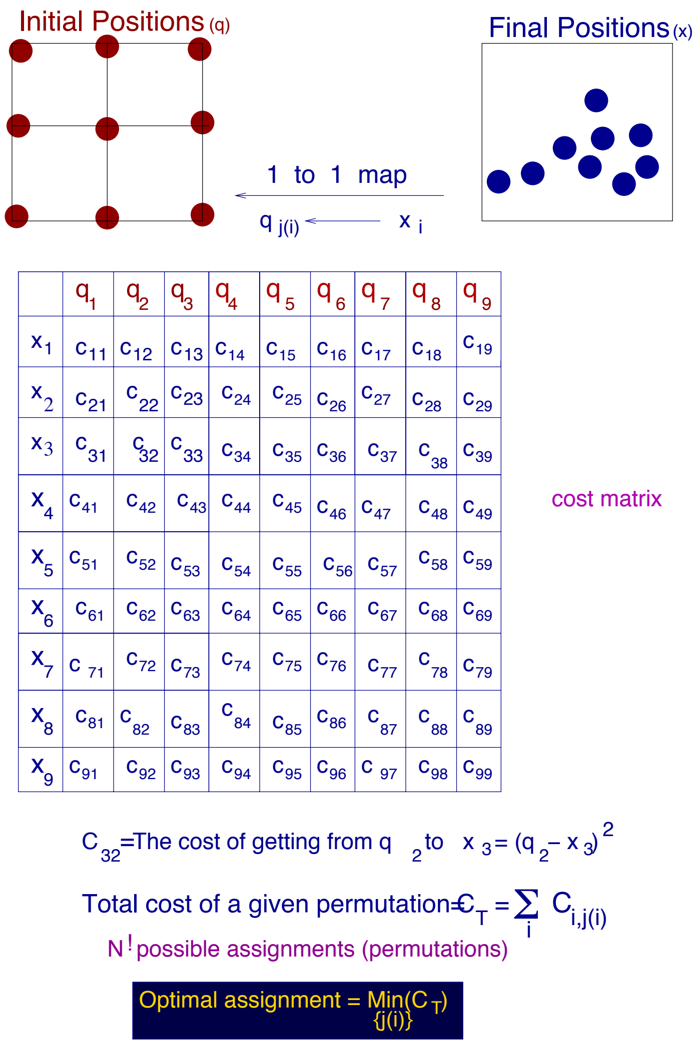

FIG. 2. Solving the reconstruction problem as an assignment problem. An example of a system of $N = 9$ is sketched. For such system there are 362,889 different costs possible, each obtained by a different permutations. Algorithms with factorial complexity are clearly impractical even for small systems. However, assignment algorithms have complexities of polynomial degrees.

IV. SOLVING THE ASSIGNMENT PROBLEM

If one were to solve the assignment problem for N particles directly, one would need to search in $N!$ possible permutations for the one which has the minimum cost. However, advanced assignment algorithms exist which reduce the complexity of the problem from factorial to polynomial (so far at best to approximately $N^{2.5}$, *e.g.*, Burkard and Derigs 1980 and Bertsekas 1998). Before discussing these methods, let us briefly comment on a class of stochastic algorithms, which do not give uniqueness and hence should be avoided.

In the PIZA method of solving the assignment problem (Croft & Gaztañaga 1997), initially a random pairing between N Lagrangian and N Eulerian coordinates is made. Starting from this initial random one-to-one assignment, subsequently a pair (corresponding to two Eulerian and two Lagrangian positions) is chosen at random. For example, let us consider the randomly-selected pair $\mathbf{x_1}$ and $\mathbf{x_2}$ which have been assigned in the initial random assignment to $\mathbf{q}_1$ and $\mathbf{q}_2$ respectively. Next one swaps their Lagrangian coordinates and assigns $\mathbf{x_1}$ to $\mathbf{q}_2$ and $\mathbf{x_2}$ to $\mathbf{q}_1$ in this example. If

$$\left[(\mathbf{x}_1 - \mathbf{q}_1)^2 + (\mathbf{x}_2 - \mathbf{q}_2)^2\right] > \left[(\mathbf{x}_1 - \mathbf{q}_2)^2 + (\mathbf{x}_2 - \mathbf{q}_1)^2\right] \tag{20}$$

then one swaps the Lagrangian positions, otherwise, one keeps the original assignment. This process is repeated until one is convinced that a lower cost cannot be achieved. However, in this manner there is no guarantee that the optimal assignment has been achieved and the true minimum cost has been found. Moreover, there is a possibility of deadlock when the cost can be decreased only by a simultaneous interchange of three or more particles, while the PIZA algorithm reports a spurious minimum of the cost with respect to two-particle interchanges. Results obtained in this way depend strongly on the choice of initial random assignment and on the random selection of the pairs and suffer severely from the lack of uniqueness (see Fig. 3).

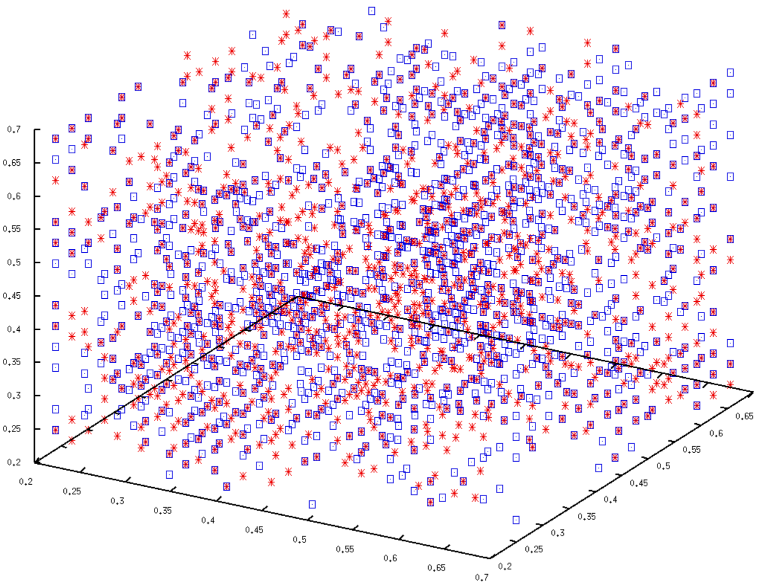

FIG. 3. The lack of uniqueness in the results of two runs using a stochastic algorithm to solve the assignment problems. In the stochastic algorithm, based on pair-interchange, one finds what is frequently referred to in pure mathematics as a *monotonic map* instead of a true *cyclic monotone map*. The red and blue points (stars and boxes respectively) are the perfectly-reconstructed Lagrangian positions using a single stochastic code with two different random seeds. The outputs of the two runs do not coincide meaning that the reconstructed Lagrangian positions (and hence peculiar velocities) depend of the initial random assignment and on the random pair interchange. The lack of uniqueness in this case is superficial and a shortcoming of the chosen numerical method. Such difficulties do not arise when deterministic algorithms are used to solve the assignment problem.

There are various deterministic algorithms which guarantee that the optimal assignment is found. An example of this is a code written by M. Hénon (Hénon 1995), demonstrated in Fig. 4. In this approach, a simple mechanical device is built which solves the assignment problem. The device acts as an *analog computer*: the numbers entering the problem are represented by physical quantities and the equations are replaced by physical laws.

The device is shown on the left side of Fig. 4. One starts with N columns Bs (which represent the Lagrangian positions of the particles) and N rows, As (which represent the Eulerian positions of the particles). On each row there are N studs whose lengths are directly related to the distances between that row and each column. The column are given weights of 1 and the rows acts as floats and are given negative weights of -1.

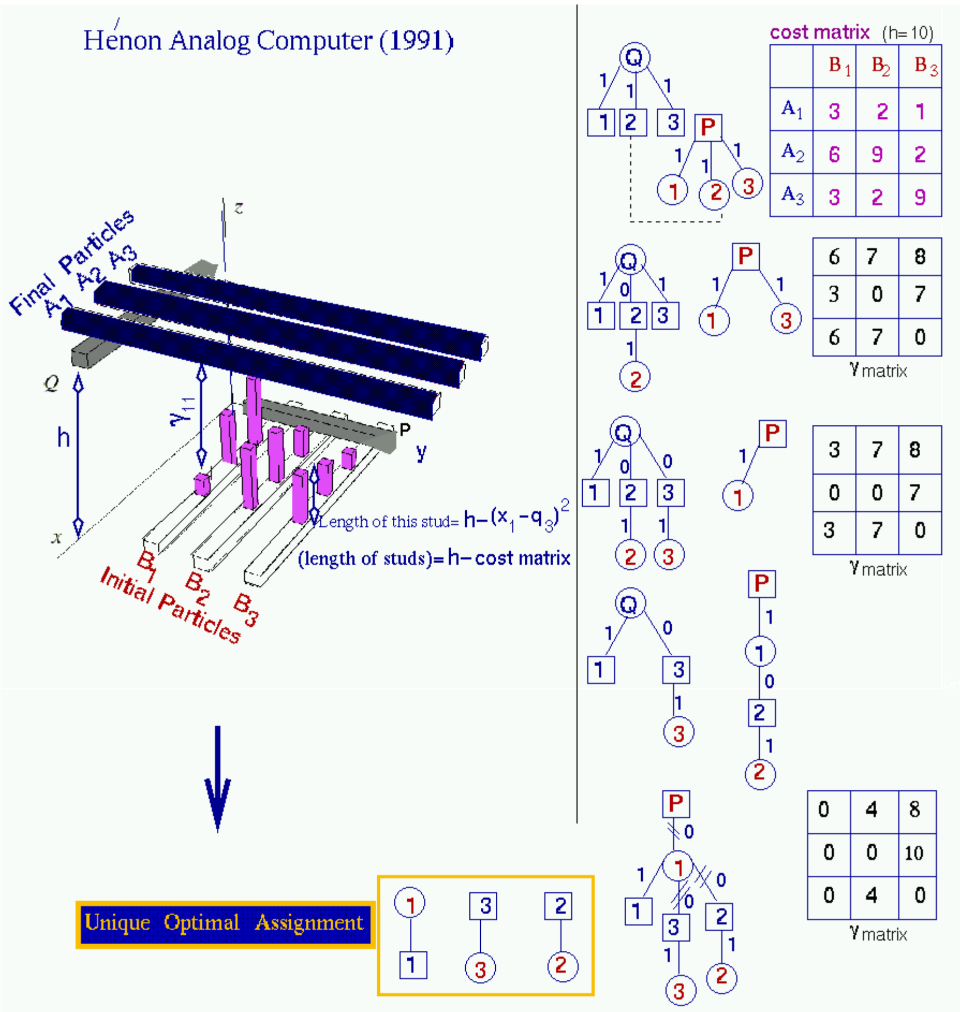

FIG. 4. Hénon's analog computer for the solution of the assignment problem is shown on the left. On the right a step-by-step progress of the algorithm on a simple example with three initial and final positions (columns and rows) is shown. The table on the top right shows the values of the costs (distances between rows and columns). When executing the algorithm by hand, it is convenient to keep track of the distances between the rows and the studs, *i.e.* the quantities $\gamma_{ij} = \alpha_i - \beta_j - c_{ij}$ where α_i is the variable height of the row A_i and β_j is the variable height of the column B_j. A graph of the initial state is made with the first contact being made between the second row and second column which have the largest cost (note that the code originally written by Hénon, in fact, finds the maximum cost. For our purpose of finding the minimum cost, one can just simply subtract the matrix elements c_{ij} from a large number). Thus, at this point the entries in the γ matrix change since now the second row and column are in contact and hence $c_{22} = 0$. Obviously the distances between all other rows and columns should also be modified. The second matrix shows the new distances and automatically a contact is made between third row and third column whose separation is now also zero. Since the second and third rows cannot move now, the next contact is made between the first column and the second row and the break occurs where the total force on the branch is weakest. Since this is where the second row meets the support Q, part of the Q tree is captured by the P tree, as demonstrated. The next contact can now only be made between the first row and the first column and the break occurs at the weakest branch. The equilibrium position is now reached where each row is supported by one column. For this simple exercise one can easily see that this procedure achieves the maximum cost (which in this example is 21).

The potential energy of this system, within an additive constant, is

$$U = \sum_i \alpha_i - \sum_j \beta_j \tag{21}$$

where α_i is the height of the row A_i and β_j is the height of the column B_j since all rows and columns have the same weight (1 and -1 respectively).

Initially, all rods are maintained at a fixed position by two additional rods P and Q with the row A_i above column B_j, so that there is no contact between the rows and the studs. Next, the rods are released by removing the holds P and Q and the system starts evolving: rows go down and columns go up and the contacts are made with the studs. Aggregates of rows and columns are progressively formed. As new contacts are made, these aggregates are modified and thus a complicated evolution follows which is graphically demonstrated with a simple example on the right side of Fig. 4. One can then show that an equilibrium where the potential energy (21) of the system is minimum will be reached after a *finite time*. It can then be shown that if the system is in equilibrium and the column B_j is in contact with row A_i then the force f_{ij} is the optimal solution of the assignment problem. The potential energy of the corresponding equilibrium is equivalent to the total *cost* of the optimal solution.

V. TEST OF MONGE-AMPÈRE-KANTOROVICH (MAK) RECONSTRUCTION WITH N-BODY SIMULATIONS

Thus, in our reconstruction method, the initial positions of the particles are uniquely found by solving the assignment problem. This result was based on our reconstruction hypothesis. We could test the validity of our hypothesis by direct comparison with numerical N-body simulations which is what we shall demonstrate later in this section. However, it is worth commenting briefly on the theoretical, observational and numerical justifications for our hypothesis. It is well-known that Zel'dovich approximation (Zel'dovich 1970) works well in describing the large-scale structure of the Universe. In the Zel'dovich approximation particles move with their initial velocities on inertial trajectories in appropriately redefined coordinates. It is also known that the gradientness of the particle displacements (expressed in Lagrangian coordinates) remains valid even at the second-order in the Lagrangian perturbation theory (Moutarde et al. 1991, Catelan 1995). This provides the theoretical motivation for our first hypothesis. The lack of multi-stream regions is confirmed by the boundedness of the cosmological structures. If multi-streaming was a significant problem in cosmology one would not observe the formation of long-lived structures such as filaments and great walls in numerical N-body simulations. In the presence of significant multi-streaming these structures would form and would smear out and disappear rapidly. This is not backed by numerical simulations which show the formation of shock-like structures well-described by Burgers models (which is a model of shock formation and evolution in a compressible fluid) which has been used to describe the large-scale structure of the Universe (Gurbatov & Saichev 1984, Shandarin & Zeldovich 1989). The success of Burgers model indicates that a viscosity-generating mechanism operates at small scales in collisionless systems resulting in the formation of shock-like structures rather than caustic-like structures.

However, in spite of this evidence in support of our reconstruction hypothesis, one needs to test it before applying it to real galaxy catalogues. We have tested our reconstruction against numerical N-body simulation. We ran a ΛCDM simulation of 128^3 dark matter particles, using the adaptive P^3M code HYDRA (Couchman et al. 1995). Our cosmological parameters are $\Omega_m = 0.3, \Omega_\lambda = 0.7, h = 0.65, \sigma_8 = 0.9$ and a box size of 200Mpc/h. We took a sample of $20,000$ and $100,000$ particles corresponding to grids of 6Mpc and 3Mpc respectively, at $z = 0$ and placed them initially on a uniform grid. An assignment algorithm, similar to that of M. Hénon described in the previous section, is used to find the correspondences between the Eulerian and the Lagrangian positions. The results of our test are shown in Fig. 5.

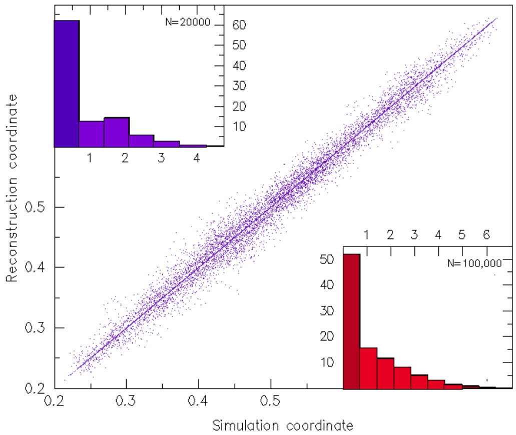

FIG. 5. Test of MAK reconstruction of the Lagrangian positions, using a ΛCDM simulation of 128^3 particles in a box of size 200 Mpc^3/h^3. In the scatter plot, the dots near the diagonal are a scatter plot of reconstructed initial points versus simulation initial points for a grid of size 6 Mpc/h with about 20,000 points. The scatter diagram uses a *quasi-periodic projection* coordinate $\tilde{\mathbf{q}} \equiv (q_x + \sqrt{2}q_y + \sqrt{3}q_z)/(1 + \sqrt{2} + \sqrt{3})$ which guarantees a one-to-one correspondence between $\tilde{\mathbf{q}}$ values and points on the regular Lagrangian grid. The upper left inset is a histogram (by percentage) of distances in reconstruction mesh units between such points; the first bin corresponds to perfect reconstruction; the lower-inset is a similar histogram for reconstruction on a finer grid of about 3 Mpc/h using 100,000 points. With the 6 Mpc/h grid 62%, and with 3 Mpc/h grid more than 50%, are assigned perfectly.

When reconstructing from observational data, in redshift space, the galaxies positions are displaced radially by an amount proportional to the radial component of the peculiar velocity in the line of sight. We have also performed another reconstruction with modified cost function which led to a somewhat degraded results but at the same time provided an approximate determination of peculiar velocities (see Fig. 6). More accurate a determination of peculiar velocities can be made using second-order Lagrangian perturbation theory. The effect of catalogue selection function and bias can be handled by giving each galaxy a *mass* which is inversely proportional to the catalogue selection function.

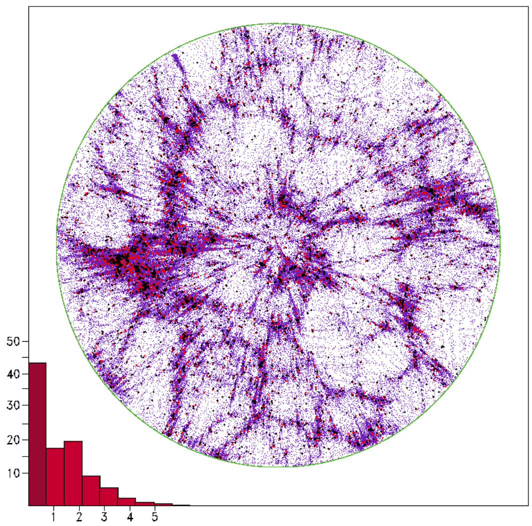

FIG. 6. Reconstruction test in redshift space with the same data as that used for real-space reconstruction tested in the upper left histogram of Fig. 5. The observer is taken to be at the centre of the simulation box. Points used for reconstruction within the displayed slice are highlighted in red. Reconstruction is performed by the MAK algorithm with a different cost function, obtained by estimating the peculiar velocities $\mathbf{v}$ using the Zel'dovich approximation: $\mathbf{v} = f(\mathbf{x} - \mathbf{q})$ where $f \approx \Omega_m^{0.6} \approx 0.49$. We have 43% of exactly reconstructed points.

In order to trace where exactly exact reconstruction is not achieved we have highlighted these points in Fig. 7. We see that exact reconstruction is not achieved in particular in the dense regions. Achieving reconstruction at small scales remains a subject of on-going research. As long as multi-streaming effects are unimportant, that is above $\simeq 1$ Mpc, uniqueness of the reconstruction is guaranteed. Approximate algorithms, capturing effects beyond Zel'dovich approximation are now being developed.

We thank M. Hénon for invaluable discussions. R.M. is supported by the European TMR network (contract HPRN-CT-2000-00162).

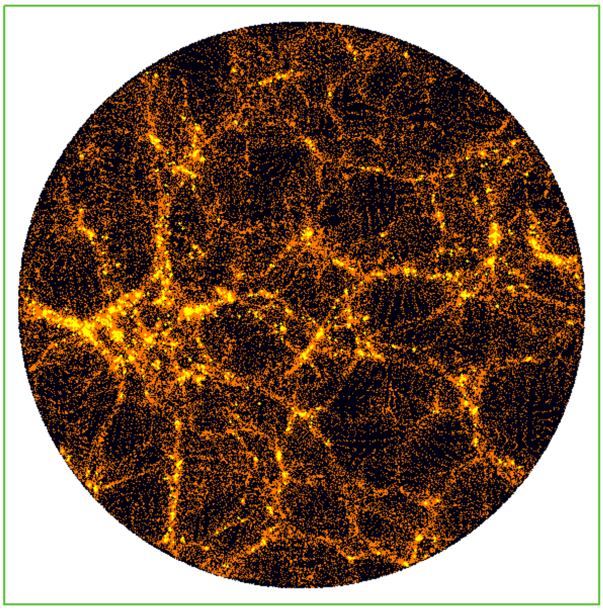

FIG. 7. N-body simulation output (present epoch) used for testing our reconstruction method. Shown here is a projection onto the $x-y$ plane of a 10% vertical slice of the simulation box of size 200 Mpc/h. the model, ΛCDM uses cold dark matter with cosmological constant and the following parameters: Hubble constant $h = 0.65, \Omega_\Lambda = 0.7, \Omega_m = 0.3$ and linear *rms* fluctuation level $\sigma_8 = 0.99$. Points are highlighted in yellow when reconstruction fails by more than 6 Mpc/h which happens mainly in the high density region where we do not expect our reconstruction hypothesis to be valid due to severe multi-streaming and generation of vorticity (in the displacements of the particles expressed in Lagrangian coordinates).

---

[1] Bertsekas D.P., *Network Optimisation: Continuous and Discrete Models* (Athena Scientific 1998)
[2] Bertschinger E. and Dekel A. 1989, ApJ **336**, L5
[3] Branchini E., Nusser A. and Eldar A. 2001, MNRAS **335**, 53
[4] Brenier Y. 1987, C.R.Acad.Sci. **305**, 805
[5] Burkard R.E. and Derigs U., *Assignment and Matching problems: Solution Methods with FORTRAN-programs*, Lecture Notes in Economics and Mathematical Systems No. 184 (Springer, Berlin 1980)
[6] Catelan P. 1995, MNRAS **276**, 115
[7] Couchman H.M.P., Thomas P.A. and Pearce F.R. 1995, ApJ **452**, 797
[8] Croft R.A. and Gaztañaga 1997, MNRAS **285**, 793
[9] Frisch U., Matarrese S., Mohayaee R. and Sobolevskii A. 2002, Nature **417**, 260
[10] Gurbatov S. and Saichev A.I. 1984, Radiophys. Quant. Electr. **27**, 303
[11] Hénon M.A. 1995, C.R.Acad.Sci. **321**, 741 A detailed version with the optimisation algorithm is available at http://arxive.org/abs/math.OC/0209047
[12] Moutarde F., Alimi J.-M., Bouchet F.R., Pellat R. and Ramani A. 1991, ApJ **382**, 377
[13] Peebles P.J.E. *The large scale structure of the Universe* (Princeton University Press, 1980)
[14] Peebles P.J.E. 1989, ApJ **L344**, 53
[15] Shandarin S.F. and Zel'dovich Y.B. 1989, Rev. Mod. Phys. **61**, 185
[16] Shaya E.J., Peebles P.J.E. and Tully R.B. 1995, ApJ **454**, 15
[17] Zel'dovich Y.B. 1970, Astron. & Astrophys. **5**, 84