Lagrangian reconstruction of cosmic velocity fields

Guilhem Lavaux\(^1,2,3,\ast\)

\(^1\)Institut d’Astrophysique de Paris, 98bis Bd Arago, 75015 PARIS, FRANCE
\(^2\)Université Paris 6
\(^3\)Université Paris 11

We discuss a Lagrangian reconstruction method of the velocity field from galaxy redshift catalog that takes its root in the Euler equation. This results in a “functional” of the velocity field which must be minimized. This is helped by an algorithm solving the minimization of cost-flow problems. The results obtained by applying this method to cosmological problems are shown and boundary effects happening in real observational cases are then discussed. Finally, a statistical model of the errors made by the reconstruction method is proposed.

PACS numbers: 47.10.A-,47.15.km,47.11.Fg,95.35.+d,98.62.Py

I. INTRODUCTION

Cosmologists are highly interested in studying galaxy peculiar velocities. Indeed, their study is a direct way to measure the dynamical state of a system and would thus permit to better understand dark matter distribution in our local Universe. The main difficulty is that measured velocities are only available sparsely and hence does not provide a good probe of the matter distribution. One must then devise an algorithm that is able to predict, under fair hypotheses, galaxy peculiar velocities from their present positions, which are their sky coordinates and their redshift, and compare the result to the measurement. Jim Peebles\(^1\) first tried to do full orbit reconstruction by evolving the present system back in time. This method proved to be quite accurate for very small volume and number of objects. However, whenever one tries to reconstruct orbits of a large number of galaxies, the method fails because the number of plausible solution is blowing up. A simplification of this problem is presented: 3D galaxy positions are assumed to be known and a simpler gravitational dynamic model is going to be assumed. We will also assume that the dynamics of galaxies is mostly driven by collisionless dark matter particles.

This proceeding is organized as follows. In §\(\Pi\) we recall the principal result of the reconstruction method developed by\(^2\) (see also the companion paper Mohayaee & Sobolevskii, hereafter MS). The method requests the using of a special fast algorithm to solve the problem. This algorithm is presented in §\(\Pi\Pi\Pi\). The method is then applied to a dark matter distribution obtained from a cosmological simulation and the reconstructed velocities are checked against the simulated ones (§\(\Pi\Pi\). Finally, a discussion on problems with bad boundary conditions, as usually met in observational cosmology, is quickly discussed in §\(\Pi\Pi\).
FIG. 1: Application to Cosmology – Top left: A slice of the density field of the ΛCDM simulation that is used for the tests (shades of gray indicate logarithm of the mass density). Top right: Adaptively smoothed line-of-sight component of the velocity field in the same slice. Bottom right: MAK reconstructed line-of-sight component of the velocity field of the same slice. Linear color scale: dark blue=-1000 km s⁻¹, white=+1000 km s⁻¹. Bottom left: Scatter plot between reconstructed and simulated velocities for objects identified in the simulation. Shades of grey show levels of the logarithm of the point density.

A. Auction algorithm

The algorithm tries to evolve the pairing map σ between \( x_i \) and \( q_j \) such that when the function is stationary between two consequent iteration it corresponds to minimizing the given total association cost. Particles locate at different Eulerian positions \( x_i \) compete against each other for Lagrangian positions \( q_j \). Minimization of the total association cost \( S_\sigma \) is achieved by studying the dual problem of minimization of association penalties \( p_j \).

In [2] it is shown that

\[
\min_{\sigma} S_\sigma = \max_{p_{j},j=1,\ldots,n} \left\{ \sum_j p_j + \sum_i r_i \right\}, \tag{3}
\]

with \( a_{i,j} = \mu(x_i - q_j)^2 \), the cost of associating \( x_i \) to \( q_j \) and \( r_i = \min_j (a_{i,j} + p_j) \). Once the set \( \{p_j\} \) is determined by the above maximization, the map \( \sigma \) is simply given by:

\[
\sigma(i) = \arg\min_j \{a_{i,j} + p_j\}. \tag{4}
\]

Effectively, \( \{p_j\} \) is computed iteratively by the algorithm. Each iteration is composed of two parts. During the first one, we obtain a set of best assignment \( A(j) \) for each particle \( q_j \) by minimizing all possible \( r_i \). Then, we link \( x_i \) to \( q_j \) with \( i^*_j \) being the particle having the minimal \( r_{i^*_j} \) in the set \( A(j) \). We also have a reverse mapping for this link that we write \( j^*_i \). Finally, the penalty \( p_j \) is updated such that

\[
p_j \to \tilde{p}_j = a_{i^*_j,j} + w_{i^*_j} - \epsilon, \tag{5}
\]

with \( \epsilon > 0 \) and

\[
w_i = \min_{j \neq j^*_i} (a_{i,j} + p_j). \tag{6}
\]

The solution found is the same as for \( \epsilon = 0 \) provided \( \epsilon < \epsilon_0/N \), with

\[
\epsilon_0 = \min_{\{i,j\} / a_{i,j} \neq 0} a_{i,j}. \tag{7}
\]

The time complexity depends quite a lot on the way \( \epsilon \) is scaled down from its initial value to the \( \epsilon_0/N \). Numerical experiments have shown that trying to converge in about 5 iterations and starting from \( \epsilon/\epsilon_0 \approx N/2 \) seems to give a faster convergence.

B. Implementation

We developed a C++ multithreaded (shared memory parallelism) and MPI version of the “Auction” algorithm, it will be available later as a multi-purpose library for cost-flow problems at the address [http://www.iap.fr/users/lavaux/](http://www.iap.fr/users/lavaux/). Besides doing a full minimization over all \( q_j \) for a given \( x_i \) ("dense" mode). It also supports a “sparse” mode that solves a partial minimization problem: for a given \( x_i \), it only minimizes over a subset of \( \{q_j\} \) such that \( ||x_i - q_j|| < R \), where \( R \) is a parameter given at the initialization to the algorithm. This allows to reduce drastically the computing time while giving the same result provided that \( R \) is not too small (typically \( R = 40 \text{ h}^{-1} \text{ Mpc} \) for a ΛCDM Universe). On a Dual-core AMD Athlon64 4800+, the SMP implementation (dense mode) takes 50 mins to assigning 79,000 particles. It has successfully reconstructed a 128³ dense mesh in a month in sparse mode. The MPI version of the corresponding algorithm is only performant for larger number of particles (typically \( N \gtrsim 500,000 \)).

Most of the time is, at the moment, spent at computing \( \min_j (a_{i,j} + p_j) \) as the cost values are only kept in a minimalistic cache. Precomputing the costs is also not feasible because of the excessive amount of memory that would be needed to store all costs for all \( (i,j) \) pairs. We also consider to implement a general purpose totally asynchronous implementation in the near future.

IV. APPLICATION TO COSMOLOGY: TEST ON COSMOLOGICAL SIMULATION

To check that the dark matter dynamical model is working, we are testing it against a 128³ \( N \)-body sample
of the algorithm described in Section 2. Finally we computed the convex map-
ning of this object. We distributed the particles of the original simulation which has been bound
object velocities, in the bottom-left panel, are exception-
ally well reconstructed. Visual inspection of the line-
tion Eq. (2), using the same cosmology as the simulation
particle were computed using the Zel’dovich approxima-
tion. A limit of 5 linked particles is put to bind particles
into a halo. The particles left unbound by this criterion
were kept in a set called the “background field”. All
objects are kept in a mock catalogue called FullMock.
We have run a reconstruction on FullMock using a MAK
mesh with $128^3$ elements. Each object of FullMock was
given a number of elements $x_i$ equal to the number of
particles of the original simulation which has been bound
into this object. We distributed the $q_j$ mesh elements
regularly on a cubic grid of the same physical size as the
simulation box. Finally we computed the convex mapping $σ$
corresponding to the MAK problem with the help of
the algorithm described in [3]. The velocities for each
particle were computed using the Zel’dovich approxima-
tion Eq. (2), using the same cosmology as the simulation
to compute $D(t)$.

Fig. 2 summarizes the results obtained using the MAK
method on the reconstructed velocities. The individual
object velocities, in the bottom-left panel, are exception-
ally well reconstructed. Visual inspection of the line-
of-sight component of the velocity field in the two right
panels show nearly no discrepancy except in regions with
really high velocities. In these regions, the dynamics is
highly non-linear, which means that the convex hypoth-
esis is not valid anymore. This problem arises on a typi-
cal cosmological scale of at most a few Mpc around large
clusters. Indeed, in those regions the fluid description of
dark matter particles completely fail because the mass
tracers may have already crossed the center of the grav-
itational attractor and are currently falling back to the
center, as illustrated Fig. 2. This renders the displace-
ment field reconstruction dubious in those cases.

V. APPLICATION TO COSMOLOGY:
BOUNDARY PROBLEMS

One does not necessarily know the Lagrangian domain $q$
on which the MAK reconstruction must be computed.
This is the case for real cosmological observations and
one must use some empirical prescription to attenuate
the boundary effects on reconstructed velocities. This
scheme is helped by the overall homogeneity of the Uni-
verse above scales larger than $200 h^{-1}$ Mpc. We propose
thus to check two schemes to handle boundary effects:

- A naive approach would be to assume that the
piece of Universe considered has not changed its
volume sufficiently between initial time and the cur-
tent time. This means that we may assume that if
we select a ball of matter, in the Universe, centered
on us, all the mass that is inside this ball is coming
from the same homogeneous ball in the Universe as
it was at decoupling time. We call this approach
NaiveDom. It is equivalent to say that tidal field
effects are totally negligible on the considered scale.

- An alternative approach is not to make an assump-
tion on the exact shape but on the low amount of
fluctuation on the boundary. Consequently, if one

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2}
\caption{Cosmology / Multi-streaming regions – This figure illustrates the different problems that may occur for a halo of dark matter particles near a cluster of galaxies. Galaxy A is in the region of first infall. The displacement field will be well reconstructed. Galaxy B is coming from the same direction as Galaxy A but has already gone through the center of the cluster and is decelerating. In that case, its displacement is badly reconstructed as, most likely, MAK predicts that the matter composing Galaxy B is coming from the region opposite to Galaxy A’s region. Galaxy C is also wrongly reconstructed.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3}
\caption{Cosmology / Boundary problems – Left panel: Illustration of the NaiveDom approach to handle boundary problems while doing a reconstruction. The dark starry ball illustrates the current dark matter distribution as inferred from galaxy catalogues. The whitish transparent ball is the assumed initial volume for the dark matter that has fallen in present structures. Right panel: Same as left panel but this illustrates the PaddedDom approach.}
\end{figure}
VI. STATISTICAL ANALYSIS OF ERRORS IN THE RECONSTRUCTION

The measurement of the slope between velocities and reconstructed displacements should give an estimation of \( \Omega_m \). However, building a reliable estimator of this slope without the statistical model of errors made both at the observation and the reconstruction level may produce unacceptable bias. We propose to show how to use models on reconstruction errors to make a bayesian analysis of the reconstructed velocities. We will focus here on errors made during a reconstruction and assume that the observed peculiar velocities \( v \) are equal to their true velocities. A more detailed discussion can be found in [6].

Using simulations, we have measured the distribution of reconstruction errors, for each object \( i \) of a catalog of galaxy, \( \{ e_i \} \) defined as

\[
e = v - \beta \psi_{r,\text{rec}} \tag{8}
\]

with \( \beta = 0.51 \) for the studied simulation (corresponding to \( \Omega_m = 0.30 \)), \( v \) the line-of-sight component of the simulated velocity of the considered, \( \psi_{r,\text{rec}} \) the reconstructed radial displacement. The result is given in Fig. 5. We have tried to fit an histogram of the errors \( \{ e_i \} \) by both a Gaussian function of width \( B \)

\[
f_G(e) \propto \exp\left( -\frac{e^2}{B^2} \right) \tag{9}
\]

and a Lorentzian function

\[
f_L(e) \propto \frac{1}{1 + \frac{e^2}{B^2}}. \tag{10}
\]

We obtained approximately the same width \( B \) for the two fits (which is expected from the second order development of both functions), however it is striking that \( f_L \) is a much better approximation than \( f_G \) to the observed error distribution.

We equate the probability of getting an error \( e \) on the true velocity \( v \) for an object of the catalog to \( f_L(e) \). We also assume now that the distribution of velocities in the object sample is, for a sufficiently large volume, Gaussian with a width \( \sigma_v \):

\[
P(v_r|\sigma_v) \propto \exp\left( -\frac{v_r^2}{2\sigma_v^2} \right) \tag{11}
\]

Now we can build the joint probability of getting \( v_r, \psi_{r,\text{rec}} \) and \( \beta \):

\[
P(v_r, \psi_{r,\text{rec}}, \beta|B, \sigma_v) \propto P(e(v_r, \psi_{r,\text{rec}})|B, \sigma_v) P(v_r|B, \sigma_v) P(\psi_{r,\text{rec}}|B, \sigma_v)
\]

\[
\propto P(\psi_{r,\text{rec}}|B, \sigma_v) \exp\left( -\frac{\psi_{r,\text{rec}}^2}{2\sigma_v^2} \right) \frac{1}{1 + \left( \frac{v_r - \beta \psi_{r,\text{rec}}}{B} \right)^2} \tag{12}
\]

where the constant of proportionality eventually depends on \( B, \sigma_v \) and \( \beta \). Using the theorem of Bayes, it is now
Using that approach we have made measurements in finite volume mock catalogs. For example, with a PaddedDom reconstruction, one measure $\Omega_m = 0.34$ with this approach (for an effective $\Omega_m = 0.35$ in this catalog), whereas a naive measurement would yield $\Omega_m \leq 0.26$.

VII. CONCLUSION

We presented a method to predict velocities of galaxies from their current position. To solve this problem, we implemented a fast algorithm invented by Dimitri Bertsekas [3] and applied the method to a pure dark matter simulation. It happens that the reconstructed velocities are impressively accurate on large-scales ([4][5]). However, the solution is only approximate in regions where multi-streaming occurs.

We proposed two methods for partially correcting boundary effects ([6][7]) and showed how boundary effects affect the reconstructed velocity field. We preferred the PaddedDom reconstruction scheme as it seems to give overall better results. Empirically we found that a buffer zone of $20 \, h^{-1} \text{Mpc}$ is needed before obtaining a reconstructed velocity field correlated with the one given by the simulation.

At last, we proposed a bayesian model ([8][9]) to account for reconstruction errors while estimating the slope between the reconstructed displacements and the true velocities of objects in a galaxy catalogs.

We would like to continue this work by improving the padding schemes to have even less boundary effects and make full use of available data in astronomy. We are also working on an improved algorithm that is able to take into account in a better way the non-linearities that are introduced in the velocity field due to gravitational effects occurring along particle trajectories. This new algorithm will try to fully solve the Euler-Poisson problem.

This work is partially supported by the ANR grant BLAN07-2_183172 (project OTARIE).

[1] P. J. E. Peebles, Tracing galaxy orbits back in time, ApJL 344 (1989) L53–L56.
[2] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, A. Sobolevskii, Reconstruction of the early Universe as a convex optimization problem, MNRAS 346 (2003) 501–524.
[3] D. P. Bertsekas, A Distributed Algorithm for the Assignment Problem, MIT Press, Cambridge, MA, 1979.
[4] R. Mohayaee, H. Mathis, S. Colombi, J. Silk, Reconstruction of primordial density fields, MNRAS 365 (2006) 939–959.
[5] H. M. P. Couchman, P. A. Thomas, F. R. Pearce, Hydra: an Adaptive-Mesh Implementation of P 3M-SPH, ApJ 452 (1995) 797–+.
[6] G. Lavaux, R. Mohayaee, S. Colombi, R. B. Tully, F. Bernardeau, J. Silk, Observational biases in Lagrangian reconstructions of cosmic velocity fields, ArXiv e-prints 707.
[7] This number is obtained for a given simulation, and particles randomly until the desired average density is obtained. The worst case of this algorithm is actually $O(N^3)$, if one makes a dense search on purely random data.
[8] G. Lavaux & G. Loeper, work in progress.