Structure formation

Perturbations in general relativity

General Relativistic Treatment

In General Relativity, we need to consider perturbations in the cosmological fluid and in the metric.

We want to find a set of equations to study the time-evolution of any perturbed quantity (for each scale), and in particular we are interested in the time-evolution of the density contrast of the matter field (the basis for structure formation).

The perturbed quantities may be written as independent modes in harmonic space and the modes evolve independently in the linear regime → no spatial evolution in the linearized equations.

The set of equations are the (perturbed and linearized) Einstein equations plus energy-momentum conservation equations.

The energy-momentum conservation equation is a continuity equation in the case of a perfect fluid. In the more complex case of a system of particles with an energy distribution evolving in the phase space, the conservation equation is the (perturbed) Boltzmann equation.

The set of equations are called the Einstein-Boltzmann equations.

Metric Perturbations

It is convenient to write the diagonal Robertson-Walker (RW) metric using conformal time τ, i.e., to factorize the scale factor:

$$ds^2 = a^2(\tau)\left[-d\tau^2 + dr^2 + f_K^2(r)\, d\Omega^2\right]$$

$$dt = a\, d\tau$$

$$\frac{d}{dt} = \frac{1}{a}\frac{d}{d\tau}$$

In matrix form and using cartesian coordinates (and in the case of flat space), it yields

$$\bar{g}_{\mu\nu} \equiv a^2 \begin{pmatrix} -1 & \mathbf{0} \\ \mathbf{0} & \delta_{ij} \end{pmatrix}$$

where δ_{ij} is the Kronecker delta, i.e., the identity matrix, and i,j (spatial indexes) run from 1 to 3.

Now, the inhomogeneities in the density field (and in other sources of gravity) produce a change in the metric.

The metric becomes inhomogeneous and, if the modifications are small, it is usually written as a perturbation to Robertson-Walker metric →

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}$$

RW is called the background metric in the inhomogeneous universe.

In general, we can perturb all 10 components of the symmetric 4x4 RW metric, and we may write a general (symmetric) 4x4 metric perturbation as:

$$\delta g_{\mu\nu} \equiv a^2 \begin{pmatrix} \boxed{\overset{\mathbf{S}}{-2\phi}} & \boxed{w_i \quad \mathbf{V}} \\ \boxed{\underset{\mathbf{V}}{w_i}} & \boxed{\underset{\mathbf{T}}{-2\psi\delta_{ij} + 2h_{ij}}} \end{pmatrix}$$

This introduces 10 new random fields (which in principle, if they are independent, are expected to introduce 10 new degrees of freedom in the metric): 1S + 3V + 6T

These 10 new quantities are:

Φ (1 scalar in the component tt) ;
w_i (a vector with 3 components ti) ;
h_{ij} (a traceless tensor with 5 components ij);
ψ (the trace of the spatial tensor of the perturbed metric)

Here, the tensor was written as a traceless tensor + trace, i.e, the types of the 10 components are now: 2S + 3V + 5T

The vector and tensor components can be further decomposed as we will see next: the SVT decomposition.

This is useful because S, V and T perturbations will evolve independently of each other, and so it will simplify the system of differential Einstein equations.

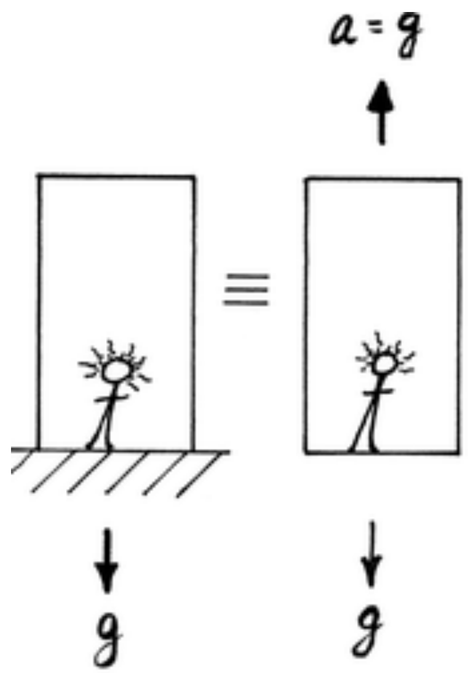
Scalar perturbations

What is the meaning of the two metric scalar perturbations?

Φ - From the **equivalence principle :**

Gravitational field (gravitational mass) ←→ Acceleration of the reference system (inertial mass)

In the well-known "gedanken" experiment in special relativity:



Photon travel time from ceiling to floor $t = h/c$
Floor's velocity increased by $g\,h/c$
Frequency shift $\Delta\nu/\nu = \Delta v/c = gh/c^2$
Time dilation $\Delta t/t = gh/c^2$
Equivalence principle → time dilation = $\Delta\Phi / c^2$

$$ds^2 = -\left(1 + \frac{2\Phi}{c^2}\right) c^2 dt^2 + dx^2$$

Minkowski metric in an accelerated frame (or with a gravitational potential: a metric perturbation)

ψ - In GR, **spatial curvature** also contributes to gravity → a perturbation to spatial curvature also changes the dynamics

$$ds^2 = - \; c^2 dt^2 + \left(1 + \frac{2\Psi}{c^2}\right) dx^2$$

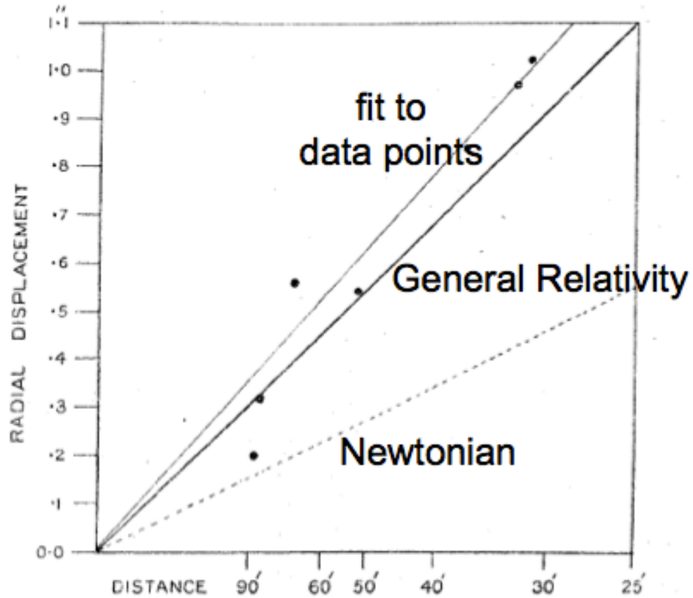
The simplest inhomogeneous metric is the scalar one - it includes two potentials, defining a **space-time curvature** that describes gravitational effects at first-order:

$$ds^2 = -\left(1 + \frac{2\Phi}{c^2}\right) c^2 dt^2 + \left(1 + \frac{2\Psi}{c^2}\right) dx^2$$

Null geodesics are determined by **Φ + ψ** (= 2Φ if they are equal),

while in Newtonian gravity, trajectories are determined by Φ (potential associated to the inertial mass), and curvature is not considered

→ **light deflection in GR is twice as large than in Newtonian gravity** → first test of GR (1919 eclipse)



Vector perturbations

Any vector can be decomposed in a sum of 2 special types of vectors:

a gradient vector
and a divergence-free vector:

$$w_i = w_{;i} + w_i^{\perp}$$

Remember ';i' stands for derivative in curved space-times , i.e., the covariant derivative with respect to the coordinate 'i'.
This is computed as the partial derivative ',i' with respect to 'i' plus the appropriate contractions with the connection.
For example the covariant derivative of a vector is: $v^a;b = v^a,b + \Gamma^a_{cb} v^c$

The decomposition leads to two important features:

i) The 3 components of the gradient vector $w_{;i}$ are all computed from derivatives of the same scalar w (this is usually called the potential of the associated vector field)
→ **the three components contain only 1 independent quantity, a scalar w**

ii) The 3 components of the divergence-free vector $w_i^\perp$ are also not all independent, they are related due to the divergence-free nature of the vector, i.e.,

$$w^{\perp 1}{}_{;1} + w^{\perp 2}{}_{;2} + w^{\perp 3}{}_{;3} = 0$$

→ the 3D divergence-free vector **has only 2 independent vector components**

So the vector perturbation with 3 components 3V was decomposed in 1S+2V

Tensor perturbations

Any traceless tensor can be decomposed in a sum of 3 special types of tensors:

a gradient of a gradient, i.e., a laplacian tensor,
a gradient of a divergence-free vector,
and a divergence-free and traceless tensor:

$$h_{ij} = D_{ij}h + h_{(i;j)} + h^T_{ij} \qquad \text{with} \quad D_{ij} \equiv \nabla_i \nabla_j - \frac{1}{3}\delta_{ij}\Delta$$

Let us look at the 3 terms:

i) In the **first term**, the tensor is $\nabla_i \nabla_j h = \begin{bmatrix} h_{;11} & h_{;12} & h_{;13} \\ h_{;21} & h_{;22} & h_{;23} \\ h_{;31} & h_{;32} & h_{;33} \end{bmatrix}$

and then the diagonal terms are subtracted by $\frac{1}{3}\left(h_{;11} + h_{;22} + h_{;33}\right)$

This results in a tensor D_{ij} h that is traceless.

Note that the full 3x3 symmetric tensor defined by the first term (containing in principle 6 independent quantities) is all built from 1 single quantity, the scalar h → **it contains only 1 independent quantity, the scalar h**

ii) The **second term** is a tensor built from a vector h_i :

$$h_{(i;j)} = \frac{1}{2}\begin{bmatrix} (h_{1;1} + h_{1:1}) & (h_{1;2} + h_{2:1}) & (h_{1;3} + h_{3:1}) \\ (h_{2;1} + h_{1:2}) & (h_{2;2} + h_{2:2}) & (h_{2;3} + h_{3:2}) \\ (h_{3;1} + h_{1:3}) & (h_{3;2} + h_{2:3}) & (h_{3;3} + h_{3:3}) \end{bmatrix}$$

so in principle it has 3 independent vector components.

However, we will consider that the vector h_i is divergence-free, i.e.

$$h_{1;1} + h_{2;2} + h_{3;3} = 0$$

This makes $h_{(i;j)}$ traceless → **the second term only contains 2 independent quantities**

iii) The **third term** is a traceless and divergence-free tensor h^T_{ij} .

This means that its 6 independent components are constrained by 4 equations:

$$h^T_{11} + h^T_{22} + h^T_{33} = 0$$

$$h^T_{11;1} + h^T_{12;2} + h^T_{13;3} = 0$$
$$h^T_{21;1} + h^T_{22;2} + h^T_{23;3} = 0$$
$$h^T_{31;1} + h^T_{32;2} + h^T_{33;3} = 0$$

→ it contains only 2 independent quantities

So the (traceless) tensor perturbation with 5 components 5T was decomposed in 1S+2V+2T

This tensor decomposition was made for a traceless tensor.

In general, the tensor perturbation do not need to be traceless. Now that the decomposition is made, it is very easy to generalize it to the case of a non-zero trace.
We just to need to add a **trace tensor**, i.e., a diagonal tensor only with the trace information → the trace is then an additional degree of freedom, and it is a scalar.

The trace tensor is usually written as $-2\,\psi\,\delta_{ij}$ → **it contains only 1 independent quantity, the trace (the scalar ψ)**

So the tensor perturbation with 6 components 6T was decomposed in 2S +2V+2T

Collecting all terms, the metric SVT perturbations are:

$$\begin{aligned}\delta g_{\mu\nu} &= \delta g^S_{\mu\nu} + \delta g^V_{\mu\nu} + \delta g^T_{\mu\nu} \\ &= a^2 \begin{pmatrix} -2\phi & w_{;i} \\ w_{;i} & -2\psi\delta_{ij} + 2h_{;ij} \end{pmatrix} + a^2 \begin{pmatrix} 0 & w_i^\perp \\ w_i^\perp & 2h_{(i;j)} \end{pmatrix} + a^2 \begin{pmatrix} 0 & 0 \\ 0 & 2h^T_{ij} \end{pmatrix}\end{aligned}$$

type	fields	constraints	degrees of freedom
scalar perturbations	ϕ,ψ,w,h	-	4
vector perturbations	$w_i^\perp,\ h_i$	$\nabla^i w_i^\perp = \nabla^i h_i = 0$	4
tensor perturbations	h^T_{ij}	$\nabla^i h^T_{ij} = (h^T)^i_i = 0,\ h^T_{ij} = h^T_{ji}$	2

So the fundamental types of the 10 degrees of freedom are:

4S + 4V + 2T

instead of 1S + 3V + 6T

Energy-Momentum Tensor Perturbations

The homogeneous metric is sourced by a perfect fluid:

$$T_{\mu\nu} = (\rho + p)\, u_\mu\, u_\nu + p\, g_{\mu\nu}$$

where u is the fluid 4-velocity $u^\mu u_\nu = -1$

A perfect fluid has no heat conduction q (a 0i vector) and no anisotropic stress π (a ij tensor). A more general fluid is:

$$T_{\mu\nu} = (\rho + p)\, u_\mu\, u_\nu + p\, g_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu + \pi_{\mu\nu}$$

The perturbed metric is sourced by a perturbed fluid:

$$T^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu = [\bar{\rho}(1+\delta) + \bar{p}(1+\delta_p)]\, u^\mu\, u_\nu + [\bar{p}(1+\delta_p)]\, g^\mu_\nu$$

$$+ [(\bar{\rho}+\bar{p})\delta u_\nu + q_\nu]\; u^\mu + [(\bar{\rho}+\bar{p})\delta u^\mu + q^\mu]\; u_\nu + \pi^\mu_\nu$$

The source quantities include components of a perturbed perfect fluid:

-- δ - density perturbation

-- δ_p - pressure perturbation

$$p = \bar{p} + \delta_p \, \bar{p}$$

The ratio of the (dimensional) pressure and density perturbations is an important property of the fluid

(as the equation-of-state that related the mean pressure and density was).

It is called the **speed of sound c_s**:

$$c_s = \left(\frac{\delta_p \, \bar{p}}{\delta \, \bar{\rho}} \right)^{1/2}$$

(it will become clear later why this quantity is the velocity of propagation of density waves in the fluid)

Pressure and density perturbations are thermodynamically related.
As the universe expands and the density decreases, pressure should also decrease if the temperature was constant (pV = nRT).
This would be an isothermal evolution. This happens in astrophysics when the process has time to thermalize (the heat transfer is fast compared with the sound speed).

However, in cosmology, the temperature of the cosmological fluid decreases with the expansion and is non-isothermal.
A special case of non-isothermal evolution is the adiabatic (or isentropic) evolution, where the temperature changes in a way that heat transfer compensates the entropy change → entropy is conserved.

Usually, pressure perturbations are separated in 2 parts: adiabatic and non-adiabatic:

$$\delta_p = \delta_{\rm pad} + \delta_{\rm pnad}$$

Now, the adiabatic case verifies $\bar{p}(\rho) = p(\bar{\rho})$

In this case, by Taylor expanding p(ρ) we can find a useful relation:

$$p(\rho) = p(\bar{\rho}) + \frac{\partial p}{\partial \rho}_{|\bar{\rho}} (\rho - \bar{\rho})$$

with $\rho = \bar{\rho} + \delta\bar{\rho}$

On the other hand, by definition

$p(\rho) = \bar{p} + \delta_p \bar{p}$ → inserting this in the Taylor expansion, it follows

$$\bar{p}(\rho) + \delta_p \bar{p} = p(\bar{\rho}) + \frac{\partial p}{\partial \rho}_{|\bar{\rho}} \delta\bar{\rho}$$

Inserting the adiabatic condition, we find that the **adiabatic speed of sound** is also given by

$$c_s = \left(\frac{\partial p}{\partial \rho}\right)_s^{1/2}$$

-- δu - velocity perturbation

In the case of the homogeneous Universe, the background 4-velocity was $u_\mu = (-a,0)$ (from its normalization) → there was no spatial velocity contribution → the homogeneous fluid was comoving with the expansion.

Now, on the contrary, there is a velocity perturbation and $u_\mu = u_\mu + \delta u_\mu$

with $\delta u_\mu = a\,(-\Phi\ , v_i + w_i)$

v_i is **the fluid velocity perturbation** - the **peculiar velocity**

w_i comes from the vector metric perturbation

The usual (SV) decomposition defines a scalar part of v_i, such that $v_i = \mathrm{grad}(\theta)$: the **scalar velocity perturbation θ** (also sometimes called v)

The source quantities may also in general include the components of a non-perfect fluid:

-- q_i - energy flux

The energy flux (it is a perturbation there is no need to define a δq, since it was zero in the background) is a velocity vector, usually decomposed in S and V parts.

-- Π_{ij} - anisotropic stress

Anisotropic perturbations in the spatial part of $T_{\mu\nu}$ form the Π_{ij}.
It is decomposed in S (written as second-order derivatives of a scalar σ), V, T

Gauge transformation

To define a metric perturbation we needs both a perturbed and an unperturbed metric → the value of the metric perturbation at (x,t) is the difference between the metric value in the inhomogeneous universe at (x,t) and the metric value that would exist without perturbations at the same point (x,t).

But it is the metric that defines the points (x,t) → the two sets of points (x,t) do not exist in the same space-time → we cannot uniquely say that one point is the same in different metrics → in order to define the perturbation we need to make an identification, a mapping, between points in the 2 metrics → the mapping fixes the gauge.

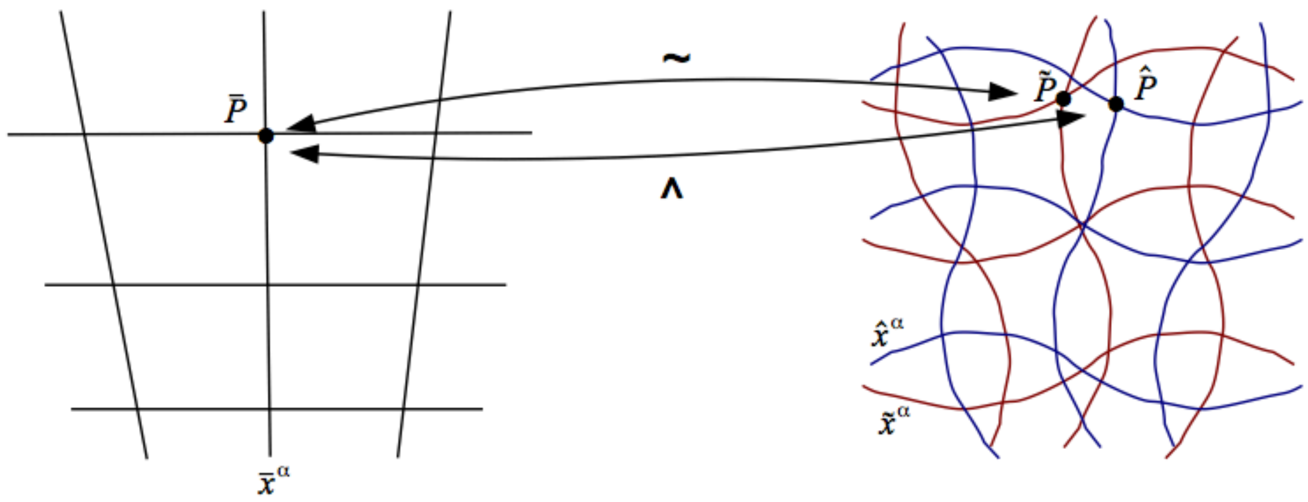
Gauge means a standard, a prescription.

Consider 2 different gauges (mappings), ~ and ^

→ a point in the background metric is identified with 2 'different' points in the perturbed metric

$$x^\alpha(\bar{P}) = \tilde{x}^\alpha(\tilde{P}) = \hat{x}^\alpha(\hat{P})$$

→ the 2 points have the same coordinates (x,t) in the 2 gauges → a quantity defined at (x,t) in both gauges may be different in the 2 gauges → it is not gauge invariant.



Consider the gauge ~

The 2 ‘equivalent’ points have different coordinates in that gauge → the transformation between the 2 points is the gauge transformation:

$$\tilde{x}^\alpha(\tilde{P}) = \tilde{x}^\alpha(\hat{P}) - \xi^\alpha$$

The transformation is described by the 4-vector $\xi^\mu = (\xi^0, \xi^i)$.

The spatial part may be decomposed as usual in a scalar and a two-component divergence-free vector:

$$\xi\,,\ \ \xi_i^\perp$$

We can apply this generic gauge transformation to any quantity defined in the space-time.

Transformation of a scalar function:

$$s(\tilde{P}) = s(\hat{P}) + \frac{\partial s}{\partial \hat{x}^\alpha} \underbrace{(\hat{x}^\alpha(\tilde{P}) - \hat{x}^\alpha(\hat{P}))}_{-\xi^\alpha}$$

(Taylor expansion)

$$\stackrel{(1)}{=} s(\hat{P}) - \frac{\partial \bar{s}}{\partial x^\alpha}\xi^\alpha \stackrel{(2)}{=} s(\hat{P}) - \bar{s}'\xi^0.$$

(1) taking the derivative in the background metric
(2) only time component is needed since the background is isotropic

(where s' is the conformal derivative of s)

In the case the **scalar is a perturbation**: $\tilde{\delta s} \equiv s(\tilde{P}) - \bar{s}(\bar{P})$

$$\tilde{\delta s} = s(\tilde{P}) - \bar{s}(\bar{P}) = s(\hat{P}) - \bar{s}(\bar{P}) - \bar{s}'\xi^0 = \hat{\delta s} - \bar{s}'\xi^0$$

The trivial solution is that any conformal time-invariant scalar is gauge-invariant.

Transformations may also be written for vectors, tensors, and for **vector perturbations**

$$\begin{aligned} \delta \tilde{w}_\mu &= w_{\tilde{\mu}}(\tilde{P}) - \bar{w}_\mu(\bar{P}) \\ &= \delta \hat{w}_\mu - \bar{w}_{\mu,\alpha}\xi^\alpha - \bar{w}_\sigma \xi^\sigma_{,\mu} \end{aligned}$$

and **tensor perturbations**

$$\begin{aligned} \delta \tilde{B}_{\mu\nu} &= B_{\tilde{\mu}\tilde{\nu}}(\tilde{P}) - \bar{B}_{\mu\nu}(\bar{P}) \\ &= \delta \hat{B}_{\mu\nu} - \bar{B}_{\mu\nu,\alpha}\xi^\alpha - \xi^\rho_{,\mu}\bar{B}_{\rho\nu} - \xi^\sigma_{,\nu}\bar{B}_{\mu\sigma} \end{aligned}$$

The total metric perturbation is a tensor perturbation, and this last formula applies.

Using the fact that the background metric is a symmetric and homogeneous tensor, we get the expression for the **gauge transformation of the metric perturbations**:

$$\delta\tilde{g}_{\mu\nu} = \delta\hat{g}_{\mu\nu} - \bar{g}'_{\mu\nu}\xi^0 - 2\xi_{(\mu,\nu)}$$

We can apply this expression to compute the gauge transformation for all metric components. For example, for the (0,0) component:

$$-2a^2\tilde{\phi} = -2a^2\hat{\phi} + 2aa'\xi^0 + 2a^2\xi^{0\prime}$$

$$\Rightarrow \tilde{\phi} = \hat{\phi} - \mathcal{H}\xi^0 - \xi^{0\prime}.$$

where $\mathcal{H}$= **a H** = a' / a

is the **conformal Hubble function**, useful when considering derivatives with respect to conformal time.

The gauge transformations of the 4 scalar components of the metric are:

$$\begin{aligned}
\tilde{\phi} &= \hat{\phi} - \mathcal{H}\xi^0 - \xi^{0\prime}, \\
\tilde{\psi} &= \hat{\psi} + \mathcal{H}\xi^0, \\
\tilde{w} &= \hat{w} + \xi^0 - \xi', \\
\tilde{h} &= \hat{h} - \xi,
\end{aligned}$$

We can also compute the gauge transformations for the **perturbed energy-momentum components**.

For example:

δ scalar perturbation: $$\tilde{\delta} = \frac{\tilde{\delta\rho}}{\bar{\rho}} = \frac{\hat{\delta\rho} - \bar{\rho}'\xi^0}{\bar{\rho}} = \hat{\delta} - \frac{\bar{\rho}'}{\bar{\rho}}\xi^0 = \hat{\delta} + 3\mathcal{H}\xi^0.$$

v_i vector perturbation: $$\tilde{v}_i^{\perp} = \hat{v}_i^{\perp} + \xi_i^{\perp\prime}$$

$$\tilde{v} = \hat{v} + \xi';$$

Fixing the gauge

Defining a particular (arbitrary) ξ fixes the gauge.

The transformation with the 4-vector ξ introduces 4 constraints between the 10 metric perturbations (2S + 2V) → reduces the number of scalar degrees of freedom from 4 to 2, and vector dof from 4 to 2 and keeps the number of tensor dof at 2 → reduces the total degrees of freedom from 10 to 6 → **there are only 6 independent components of the metric perturbations.**

Instead of defining the quadrivector ξ, the gauge can alternatively be fixed by assigning the values of 4 perturbations (2S + 2V).

Some examples of gauges:

Synchronous gauge

$w = 0$ → no cross terms x,t in the metric → allows to define comoving observers for which x does not change as time goes by (just like it happens for the background RW).

$\Phi = 0$ → all comoving observers (at different x positions) have synchronous time → no gravitational redshift (i.e., no conformal cosmological redshift).

In this gauge the two scalar perturbations remaining are:

ψ and h, that only affect the spatial ii and ij components.

(and there are also 2V and 2T d-o-f remaining)

(Conformal) Newtonian gauge (also called longitudinal gauge)

$w = 0$ → no cross terms x,t in the metric → allows to define comoving observers for which x does not change as time goes by (just like it happens for the background RW).

$h = 0$ → spatial perturbations are diagonal → no shear perturbations

In this gauge the metric is defined by ψ and Φ (besides 2V and 2T dof)

Φ gives the gravitational redshift → it is a gravitational potential (taking the limit of GR for weak fields, like in Newtonian gravity, hence the name of this gauge)

ψ is called the curvature potential

From the choice of metric perturbation values in two particular gauges, we can compute the transformation ξ between the two gauges.

For example, the scalar part of the gauge transformation between the synchronous and Newtonian gauge is:

$$\xi^0 = \frac{h'_S + \psi'_S}{2k^2}$$

(Note that here we are working with quantities in Fourier space → the transformation is function of scale.

With this we can compute the **transformations between these gauges** for all quantities:

the **density contrast** transforms between the synchronous and the Newtonian gauge as:

$$\delta_S = \delta_N - \xi^0 \frac{\bar{\rho}'}{\bar{\rho}}$$

the **velocity perturbation** transforms as: v_S = v_N + ξ'

Gauge invariance

We see that the metric perturbations are different from gauge to gauge and δ depends on the gauge,

however observables should be gauge-independent.

The measurement of a power spectrum (or δ value) should not depend on the theoretical choice of the gauge.

Looking at the gauge transformation expression, we see this is indeed the case for large values of k (small scales), and for Universes with metric perturbations that vary slowly → this is the case of sub-Hubble scales (small, intermediate and even large scales in the late universe)

Only for the 'very relativistic universe' is there an ambiguity in observations, i.e. for very large scales or in the early universe.

However, some gauge-invariant combinations can be defined. On very large scales these quantities are the ones that have physical meaning.

An example of gauge-invariant metric quantities are the Bardeen potentials:

$$\Phi \equiv \phi + \frac{1}{a}\left[(w - h')a\right]'$$

$$\Psi \equiv \psi - \mathcal{H}(w - h')$$

Notice that in the Newtonian gauge, these gauge-invariant Bardeen potentials are identical to the scalar perturbations.

An example of a gauge-invariant metric-source quantity is the curvature perturbation:

$$\zeta \equiv \frac{1}{3}\delta + \psi$$

The power spectrum of the curvature perturbation is computed in inflation → it gives the initial condition for the potential power spectrum → and consequently for the matter power spectrum.

An example of a gauge-invariant source quantity is the comoving-gauge density contrast:

$$\Delta = \delta + 3\mathcal{H}v$$