Large-Scale Structure Observables in General Relativity

Galaxy Clustering, Cosmic Rulers, and Local Tides

written by ChatGPT and Claude, supervised by Donghui Jeong (PSU/KIAS)

30 July 2026

Table of contents

Preface

These lecture notes develop the relation between the matter perturbations of a cosmological model and the numbers that appear in a galaxy catalogue. They are written for graduate students in theoretical physics who are comfortable with actions, symmetries, perturbation theory, and general covariance, but who have had less occasion to ask how a survey turns photons into a statement about spacetime and matter.

That translation is not a secondary data-analysis step. It is part of the theoretical prediction, and it is the subject of the five lectures collected here.

Why relativistic observables matter

A cosmological model is confronted with observation only after it passes through a sequence of maps,

$$\mathcal{S}[g_{\mu\nu}, \Phi_A, ...] \longrightarrow \{g_{\mu\nu}, \Phi_A\} \longrightarrow \{n_g, u_g^\mu, S_{ij}, ...\} \longrightarrow \{\tilde{z}, \hat{\mathbf{n}}, F, \gamma, N, ...\} \longrightarrow \text{statistical inference.}$$

The first arrow is dynamics: an action and an initial state determine spacetime and matter fields. The second is source physics: galaxies and other tracers form and respond to their long-wavelength environment. The third is the relativistic observation map: photons propagate from the source to an observer, who records redshifts, directions, fluxes, shapes, and counts. Only the final objects are entries in a catalog. A proposal for new dark-sector physics or modified gravity can affect several arrows at once, and its observational meaning depends on keeping them distinct.

Figure placeholder. The chain from theory to data. A single horizontal schematic showing action, source formation, photon propagation along the past light cone, observer, and catalogue, with the intermediate variables of each stage labelled beneath the corresponding arrow.

Diffeomorphism invariance in operational form

In a field-theory calculation, gauge redundancy is often handled by choosing a convenient gauge and working with the resulting fields. In cosmology, that is not the end of the problem. A telescope does not report the matter-density contrast at a coordinate position. It reports a source at an observed redshift and direction, at a specified observer proper time. Under a diffeomorphism, the coordinate representation of the density, the metric perturbations, and the coordinate position assigned to the emission event all change. The complete observable does not.

The cancellation is not an additional prescription imposed after the calculation; it follows from perturbing a single covariantly defined measurement. This is the technical thread that runs through these notes. In Chapter 2 it appears as the pullback of a current three-form; in

Chapter 3 as the ratio of a physical ruler to an inferred one; in Chapter 4 as the statement that a freely falling observer can remove the connection but not the curvature.

> **! Key idea: define the measurement before perturbing it**
>
> A coordinate-dependent field is not automatically unphysical, and a gauge-invariant combination is not automatically an observable. Define the measurement covariantly, specify the source and observer conditions, and only then perturb the result.

Why the dark universe requires observables

The dark universe is inferred through a logic of consistency. Dark matter, late-time acceleration, and primordial perturbations are not observed as labels attached to individual events. They enter a spacetime model that must account simultaneously for expansion, growth, peculiar velocities, clustering, lensing, and other measured maps. The strength of the standard cosmological picture is that these different observations can be organized into one consistent description.

Its phenomenological success does not by itself identify the microscopic nature of the dark sector. Understanding how the consistent picture is built is therefore a prerequisite for challenging it. A model that reproduces one observable while quietly changing the meaning of another has not been tested; it has been fitted.

Why this concerns theorists outside cosmology

A new degree of freedom, interaction, symmetry, or modification of gravity is not yet a cosmological prediction when its effect has only been computed for a gauge-dependent field. One must also determine how the theory changes tracer formation, local tidal response, photon propagation, and the actual variables used by an observer.

Conversely, the tools used here are familiar theoretical tools placed in an observational setting: functional response and effective field theory for galaxy bias in Chapter 1, differential forms and pullbacks for number counts in Chapter 2, gauge symmetry for cosmic clocks and rulers in Chapter 3, and the exponential map and local curvature for Fermi normal coordinates in Chapter 4. A reader who has met these objects in another context will find the cosmological application unusual mainly in what it demands of them: they must all remain consistent inside a single measured number.

Large-scale structure as an application of general relativity

In this operational sense, large-scale-structure observables are close to an endgame of general relativity. The difficulty is not only solving the field equations. Covariance, gauge freedom, null propagation, the source and observer split, local freely falling frames, and matter response must all remain consistent in a single prediction.

Three failure modes recur, and each is treated explicitly in what follows. A coordinate artifact can resemble a physical large-scale effect. A genuine physical effect can move between intermediate variables under a gauge transformation. A locally measurable gravitational field begins only at tidal order. The output of the calculation is not a metric perturbation, but a relation among quantities that an observer can measure.

The dark universe requires ideas beyond a phenomenological parameterization. The purpose of these lectures is to provide the translation layer that allows such ideas to become controlled predictions for data.

Course map

Scope

These notes develop the relation between matter perturbations, biased tracers, and observables on the observer's past light cone. The discussion is organized around three constructions: tracer bias as a functional response to long-wavelength gravitational fields, number counts as the pullback of the galaxy-current three-form to observed coordinates, and local gravitational measurements through the Manasse–Misner construction of Fermi normal coordinates. At linear order, cosmic clocks and cosmic rulers convert a density defined on a source proper-time hypersurface into an observed number-count field. The same local-curvature viewpoint then relates the Newtonian tidal operator, intrinsic alignment, and the response to gravitational waves.

The main references are Secs. 2.5 and 9.3 of[1],[2],[3],[4], and[5].

The central distinction is

$$\delta_m\big|_{t_F} \xrightarrow{\text{formation response}} \delta_g\big|_{t_F} \xrightarrow{\text{clock, ruler, and pullback}} \delta_g^{\rm obs}(\tilde{z}, \hat{\mathbf{n}}).$$

The first map describes how a selected tracer population responds to its long-wavelength environment. The second map describes how source fields and spacetime geometry produce a field on the observed redshift and angle manifold. Local gravitational effects will be expressed in terms of curvature measured in a freely falling orthonormal frame.

The distinction in this map is not only notation. The three fields answer different questions. The matter density contrast describes the state of the matter fluid on a specified spacetime slicing. The galaxy density contrast describes a selected population on a specified source hypersurface. The observed field is the number assigned to a cell in measured redshift and direction at a fixed observer proper time. They coincide only after taking a limit in which the relevant time slicings and coordinate maps can be identified.

A useful rule throughout is to define the geometric or operational object before perturbing it. A gauge transformation then changes the coordinate representation of the source field and the coordinate location of the observed source, but not the full observable. This viewpoint avoids treating gauge invariance as a cancellation to be guessed after a calculation; the cancellation follows from expanding one covariant construction.

The discussion separates two notions of observability. Number counts and ruler distortions are nonlocal observables on the past light cone: they contain source, propagation, and observer contributions. Tidal fields are local observables: after transforming to a freely falling frame, the connection can be removed on the observer's worldline, while curvature remains. Both notions are needed for large-scale-structure observables, and the division of these notes into two parts follows this split.

How to read these notes

The chapter Notation and conventions collects the assumptions, signature, and symbol conventions used throughout; readers who prefer to start with physics can skip it and return when a symbol is unfamiliar.

Part I builds the observed number count. Chapter 1 sets the Newtonian baseline and defines bias as a response. Chapter 2 replaces Euclidean number conservation by a covariant pullback. Chapter 3 assembles clocks, rulers, and the source density into the central result.

Part II turns to what a freely falling observer can measure locally. Chapter 4 constructs Fermi normal coordinates. Chapter 5 applies the resulting local curvature to tidal bias, intrinsic alignment, and gravitational waves.

The appendices are reference material rather than a continuation of the argument. They collect response conventions (Appendix A), pullback and Lie-derivative identities (Appendix B), a gauge-transformation table (Appendix C), and the index-order algebra behind the Manasse–Misner metric (Appendix D).

Four kinds of highlighted box appear in the text, and each has a fixed meaning:

Box	Meaning
Key idea	A conceptual anchor. The physics that survives when the algebra is forgotten.
Historical note	Where a construction came from, in two or three sentences.
Tip	An intuition or a mnemonic that is useful but not load bearing.
Caution	A convention trap. Ignoring one of these produces a wrong sign or a double count.

Passages marked as a figure placeholder indicate where an illustration would help. The artwork has not yet been drawn.

Notation and conventions

This chapter collects the assumptions and symbols used throughout these notes. It is reference material; nothing here is derived, and a reader starting from Chapter 1 can return to it as needed.

Standing assumptions

Unless stated otherwise:

- the background is a spatially flat FLRW spacetime;
- matter is pressureless and follows geodesics;
- metric and light-cone observables are treated to linear order in Chapter 2 and Chapter 3;
- the bias discussion assumes Gaussian, adiabatic initial conditions and general relativity;
- the metric signature is $(-,+,+,+)$;
- η is conformal time, a prime is ∂_η, $\mathcal{H} = a'/a$, and $H = \mathcal{H}/a$;
- a tilde denotes a quantity inferred directly from observed redshift and direction using the background relation, for example $\tilde{a} = (1+\tilde{z})^{-1}$ and $\tilde{\mathbf{x}} = \tilde{\chi}\hat{\mathbf{n}}$; it never denotes a gauge-transformed field, for which we write $Q \mapsto Q + \Delta_\xi Q$;
- δ_g denotes a rest-frame galaxy density perturbation, while Δ_g or $\delta_g^{\rm obs}$ denotes an observed number-count perturbation.

We use $\bar{n}_g$ for the mean physical number density and define

$$b_e \equiv \frac{\mathrm{d}\ln(a^3\bar{n}_g)}{\mathrm{d}\ln a}.$$

Thus $b_e = 0$ for a conserved comoving population.

Two conventions that are easy to trip over

Sign of the spatial potential

The conformal-Newtonian metric in Eq. 2.12 is written with $g_{ij} = a^2(1+2\Phi)\delta_{ij}$. This is the opposite sign to the widespread convention $g_{ij} = a^2(1-2\Phi_{\rm common})\delta_{ij}$. The local weak-field metric of Eq. 5.1 uses the common local convention instead, because that subsection works in local physical coordinates rather than in the conformal chart. Both choices are stated where they are used.

> **Curvature sign and index order**
>
> The Riemann convention is fixed in Eq. 4.1. The Manasse–Misner paper places the first covariant index before the contravariant one, which flips the sign of the all-lowered tensor relative to ours. Appendix D carries out the translation explicitly.

Symbols

The tables below list where each symbol is defined rather than restating its meaning in full.

Background and kinematics

Symbol	Meaning
a, η, τ	scale factor, conformal time, proper time
$\mathcal{H}$, H	conformal and physical Hubble rates, $H = \mathcal{H}/a$
χ, $\tilde{\chi}$	comoving distance, and the distance inferred from $\tilde{z}$
$D(a)$, f	linear growth factor and growth rate, Eq. 1.2
$\Omega_m(a)$	matter density parameter at scale factor a
$\mathbf{v}$, θ	peculiar velocity and its divergence, Eq. 1.3
$\hat{\mathbf{n}}$, μ	observed line of sight, and $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}$

Metric perturbations and gauge

Symbol	Meaning
A, B, D, E	scalar metric perturbations, Eq. 2.7 and Eq. 2.8
Ψ, Φ	conformal-Newtonian potentials, Eq. 2.12
h_{ij}	spatial metric perturbation; $h_\parallel = \hat{n}^i \hat{n}^j h_{ij}$
$\xi^\mu = (T, \partial^i L)$	scalar gauge generator, Eq. 2.2
Δx^μ	displacement of the true source event from the inferred one, Eq. 2.22
$\Delta \ln a$	scale-factor perturbation at fixed observed redshift, Eq. 2.24
I_A	integrated lapse along the source worldline, Eq. 2.31

Appendix C tabulates how each of these transforms.

Tracers and bias

Symbol	Meaning
n_g, $\bar{n}_g$	physical number density of the sample and its mean
δ_g, δ_g^{det}, ϵ	tracer perturbation, its deterministic part, and the stochastic residual, Eq. 1.5 to Eq. 1.8
$\mathcal{O}_L^A$	basis of long-wavelength local gravitational operators
$\mathcal{R}^{(N)}$	functional response kernel, Eq. 1.9
b_1, b_2, b_{K^2}, $b_{\nabla^2\delta}$	bias coefficients, Eq. 1.19
b_e	evolution bias, defined above
s	magnification slope of a threshold sample, Eq. 3.41
R_*, k_*	formation scale and its inverse, Eq. 1.14
K_{ij}	dimensionless trace-free tidal field, Eq. 1.18

Response normalizations are compared in Appendix A.

Counts on the light cone

Symbol	Meaning
j^μ, u_g^μ	galaxy number current and four-velocity, Eq. 2.45
ε	spacetime volume form, Eq. 2.44
$\mathcal{J}$	current three-form, Eq. 2.46
X, X^*	light-cone embedding and its pullback, Eq. 2.49
$\widetilde{\boldsymbol{\omega}}$	apparent volume form of an observed cell, Eq. 2.66
δ_g^{obs}	observed number-count perturbation, Eq. 2.69
δ_g^{or}, δ_g^{pt}	source density at the observed-redshift and proper-time references, Eq. 3.29 and Eq. 3.35

Pullback identities are collected in Appendix B.

Clocks and rulers

Symbol	Meaning
$\mathcal{T}$	cosmic clock, Eq. 2.37
P_{ij}	screen projector, Eq. 3.1
$\mathcal{C}$	longitudinal ruler distortion, Eq. 3.16
$\mathcal{B}_i$	mixed ruler distortion, Eq. 3.17
$\mathcal{A}_{ij}$	transverse ruler distortion, Eq. 3.18
$\mathcal{M}$	magnification, the trace of $\mathcal{A}_{ij}$, Eq. 3.22
$\gamma_{ij}^{\mathrm{ruler}}$	ruler shear, Eq. 3.23

Symbol	Meaning
$\hat{\kappa}$	coordinate convergence, Eq. 3.21

Local frames and tides

Symbol	Meaning
$e_{\hat{a}}$	parallel-transported orthonormal tetrad, Eq. 4.13
$X^{\hat{a}} = (T, X^i)$	Fermi normal coordinates, Eq. 4.17
$\mathcal{G}$	the central geodesic
$R^{\rho}{}_{\sigma\mu\nu}$	Riemann tensor in the convention of Eq. 4.1
$\mathcal{R}_{ij}$	electric part of the Riemann tensor, Eq. 4.52
$\mathcal{E}^{S}_{ij}$, $\mathcal{E}^{\mathrm{GW}}_{ij}$	scalar and tensor trace-free tides, Eq. 5.7 and Eq. 5.23
S^{I}_{ij}, γ^{I}_{ab}	intrinsic shape and its screen projection, Eq. 5.12 and Eq. 5.14
R_E	temporal shape-response kernel, Eq. 5.11
h^{TT}_{ij}	transverse-traceless metric perturbation, Eq. 5.15

Index conventions

Greek indices $\mu, \nu, \ldots$ run over spacetime and Latin indices $i, j, \ldots$ over space. Indices a, b on screen-projected quantities run over the two directions transverse to the line of sight.

A hat marks a component referred to the orthonormal tetrad carried by a freely falling observer, equivalently to the Fermi normal coordinates built from that tetrad, as used throughout Chapter 4 and Chapter 5. Frame directions are $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ for all four and $\hat{0}$ for the time direction, which is never written bare. Latin i, j, k, l, m are the three spatial frame directions; these are frame indices by definition, so the hat would carry no information and is not written. An unhatted Greek index means an arbitrary chart on spacetime and nothing more, with μ' available for a second chart.

The rule has one visible consequence worth anticipating. Chapter 5 writes R_{0i0j} with a bare time index where it works inside the weak-field chart of Eq. 5.1 or the transverse-traceless chart of Eq. 5.15, and $R_{\hat{0}i\hat{0}j}$ where the same curvature is referred to the observer's frame. Reading the hat is enough to tell which is meant.

Angle brackets denote the symmetric trace-free part, $\mathcal{R}_{\langle ij \rangle}$ as in Eq. 4.53, and round brackets on a pair of indices denote symmetrization. Square brackets around an operator, as in $[\delta^2]$ of Eq. 1.19, denote a renormalized composite operator rather than a commutator.

A bold symbol is a spatial three-vector, $\mathbf{x}$ or $\mathbf{k}$, while the same letter with an index is its component. A bold upright symbol is a differential form, as in $\boldsymbol{\mathcal{J}}$ and $\boldsymbol{\varepsilon}$.

Part I

Part I: From Matter to Observables

1 Newtonian galaxy clustering: bias as a response and redshift-space distortions

Where we are

The Preface described a chain of maps running from an action to a catalogue. This chapter takes the last two links of that chain and computes them in the approximation where the subtleties of general relativity are invisible: scales well inside the horizon, weak fields, and a distant observer. Nothing here requires the machinery of the two chapters that follow, and that is precisely why it comes first. Working the problem once without covariance shows which of its features are physics and which were assumptions hidden by the notation.

The Newtonian calculation provides the baseline against which the relativistic construction will be compared. It already contains two logically distinct maps. First, gravitational evolution produces a matter field and galaxy formation converts that field into a tracer field. Second, the radial velocity converts the tracer field in real space into a field in redshift space. The first map is encoded by bias coefficients; the second is a coordinate Jacobian. Keeping them separate will make the general-relativistic extension transparent.

Two results carry forward. The first is a definition rather than a formula: galaxy bias is the functional response of a selected population to its long-wavelength environment, and the familiar coefficients b_1, b_2, b_{K^2} are derivatives of that response. Writing bias this way, rather than as a polynomial ansatz, is what allows the coefficients to survive the move to a relativistic setting, because a response can be defined at fixed local proper time whereas a polynomial in a coordinate-dependent field cannot. The second is the Kaiser formula, which shows that anisotropy in a measured power spectrum can be manufactured entirely by the choice of radial coordinate.

By the end of the chapter we will have an explicit list of what the Newtonian derivation left unstated. That list is the agenda for Chapter 2 and Chapter 3.

Figure placeholder. The two maps of this chapter, drawn as a single left-to-right diagram: matter density on a constant-time slice, then the tracer field produced by the formation response, then the redshift-space field produced by the radial coordinate Jacobian, with the perturbative object responsible for each arrow labelled.

1.1 Matter as a Newtonian fluid

We work in comoving coordinates $\mathbf{x}$, with peculiar velocity $\mathbf{v} = \mathrm{d}\mathbf{x}/\mathrm{d}\eta$. The density contrast $\delta = (\rho_m - \bar{\rho}_m)/\bar{\rho}_m$ is defined on a constant-conformal-time hypersurface. On scales well inside the horizon and for nonrelativistic matter, the continuity, Euler, and Poisson equations close the evolution of δ, $\mathbf{v}$, and the Newtonian potential Φ:

In comoving coordinates, pressureless matter obeys

$$
\begin{aligned}
\delta' + \boldsymbol{\nabla}\cdot\left[(1+\delta)\mathbf{v}\right] &= 0, \\
\mathbf{v}' + \mathcal{H}\mathbf{v} + (\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{v} &= -\boldsymbol{\nabla}\Phi, \\
\nabla^2\Phi = 4\pi G a^2 \bar{\rho}_m \delta &= \frac{3}{2}\mathcal{H}^2\Omega_m(a)\delta.
\end{aligned}
\tag{1.1}
$$

Here $\Omega_m = \Omega_m(a)$. The continuity equation is number conservation for matter elements, the Euler equation is their geodesic equation in the weak-field limit, and the Poisson equation relates the local gravitational curvature to the density source. The nonlinear terms transport density and velocity along the flow. Dropping products of perturbations gives a scale-independent growing mode for pressureless matter in standard general relativity. At linear order,

$$
\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2\Omega_m(a)\delta = 0, \qquad \delta(\mathbf{k},a) = D(a)\delta(\mathbf{k},a_{\rm ini}), \qquad f \equiv \frac{\mathrm{d}\ln D}{\mathrm{d}\ln a}. \tag{1.2}
$$

With

$$
\delta(\mathbf{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{k}),
$$

the linear continuity equation gives

$$
v_i(\mathbf{k}) = \mathrm{i}\mathcal{H}f\frac{k_i}{k^2}\delta(\mathbf{k}), \qquad \theta \equiv \boldsymbol{\nabla}\cdot\mathbf{v} = -\mathcal{H}f\delta. \tag{1.3}
$$

The velocity relation follows from $\delta' + \theta = 0$ and $\delta' = \mathcal{H}f\delta$. It contains one inverse power of k relative to the density. Consequently a line-of-sight velocity itself is important on large scales, while a velocity gradient is of the same perturbative order as the density. This distinction will reappear when relativistic redshift terms are organized by powers of $\mathcal{H}/k$. The equal-time matter power spectrum is

$$
\langle\delta(\mathbf{k})\delta(\mathbf{k}')\rangle = (2\pi)^3\delta_D(\mathbf{k}+\mathbf{k}')P_{mm}(k). \tag{1.4}
$$

This is a theoretical field on a chosen time slice. A galaxy catalogue measures neither this field nor this time slice directly.

The power spectrum in Eq. 1.4 is therefore not yet an observable galaxy power spectrum. Even before accounting for light propagation, the objects in a catalogue are selected tracers rather than randomly chosen matter particles. Their abundance can retain information about the long-wavelength environment in which they formed.

1.2 The tracer abundance as a functional of long-wavelength fields

Let $n_g(\mathbf{x},\eta)$ be the physical number density of a chosen tracer sample and let $\bar{n}_g(\eta)$ be its ensemble mean. We define

$$
1 + \delta_g(\mathbf{x},\eta) \equiv \frac{n_g(\mathbf{x},\eta)}{\bar{n}_g(\eta)}. \tag{1.5}
$$

The definition of the sample is part of the definition of n_g: mass threshold, luminosity cut, color selection, and any other source property must be held fixed when a response is taken.

Two samples occupying the same matter distribution need not have the same bias coefficients. The statement $\delta_g = b_1\delta$ is not a microscopic identity. It is the leading term in the response of the statistically averaged tracer abundance to a slowly varying gravitational environment.

! Key idea: bias is a response, not an ansatz

It is tempting to read $\delta_g = b_1\delta + \frac{1}{2}b_2\delta^2 + \cdots$ as a Taylor series of one field in another, fitted after the fact. It is not. Each coefficient is a derivative of the conditional mean abundance with respect to a long-wavelength gravitational operator, holding the physical definition of the sample fixed. The distinction has consequences that appear immediately: a response can be evaluated in any frame in which the long-wavelength environment is specified, so it survives the passage to a relativistic calculation, whereas a polynomial in a coordinate-dependent density does not. It also tells us which coefficients can exist at all, since the allowed operators are fixed by symmetry rather than by convenience.

Split the initial conditions, or equivalently the evolved fields, into long and short modes. Denote by $\mathcal{O}_L^A$ a complete set of long-wavelength local gravitational observables; the label A distinguishes density, tidal tensors, their convective derivatives, and higher-derivative operators. Holding the long fields fixed and averaging over short modes defines a conditional mean

$$\bar{n}_g[\mathcal{O}_L](x) \equiv \big\langle n_g(x) \big\rangle_{S\,|\,\mathcal{O}_L}. \tag{1.6}$$

The conditional average does not assume that individual galaxies form deterministically. It integrates over the short modes and over unresolved formation histories while retaining the specified long-wavelength background. The result is the part of the abundance that is predictable from the long fields. Fluctuations around this conditional mean are stochastic from the viewpoint of the long-wavelength theory. The deterministic tracer response is the functional

$$\delta_g^{\rm det}[\mathcal{O}_L](x) \equiv \frac{\bar{n}_g[\mathcal{O}_L](x) - \bar{n}_g(x^0)}{\bar{n}_g(x^0)}. \tag{1.7}$$

The residual

$$\epsilon(x) \equiv \delta_g(x) - \delta_g^{\rm det}[\mathcal{O}_L](x) \tag{1.8}$$

is stochastic with respect to the long fields. Its correlations are analytic in k^2 on scales much larger than the tracer formation scale, subject to number- or momentum-conservation constraints appropriate to the tracer.

The functional Taylor coefficients are the response kernels

$$\mathcal{R}^{(N)}_{A_1\cdots A_N}(x; x_1, \dots, x_N) \equiv \frac{1}{\bar{n}_g(x^0)}\, \frac{\delta^N \bar{n}_g[\mathcal{O}_L](x)}{\delta\mathcal{O}_L^{A_1}(x_1)\cdots\delta\mathcal{O}_L^{A_N}(x_N)}\bigg|_{\mathcal{O}_L=0}. \tag{1.9}$$

The kernel of order N answers a specific question: if the long-wavelength operator $\mathcal{O}_L^{A_1}$ is perturbed at the spacetime point x_1, and $\mathcal{O}_L^{A_2}$ at x_2, and so on, by how much does the mean abundance at x change? Because the perturbation is applied at arbitrary points, the kernel is a distribution over the whole past of x rather than a number. Appendix A records the alternative normalization in which the derivatives are taken of $\ln \bar{n}_g$ instead, which agrees with this one at first order and differs beyond it. Assembling the kernels into a functional Taylor series gives

$$\delta_g^{\rm det}(x) = \sum_{N=1}^{\infty} \frac{1}{N!} \int \mathrm{d}^4 x_1 \cdots \mathrm{d}^4 x_N \, \mathcal{R}^{(N)}_{A_1 \cdots A_N}(x; x_1, \dots, x_N) \prod_{a=1}^{N} \mathcal{O}_L^{A_a}(x_a). \tag{1.10}$$

Eq. 1.9 is a definition. The bias expansion follows after imposing causality, the equivalence principle, rotational invariance, and locality on the response kernels.

At this stage the response kernel is allowed to connect the tracer at x to long fields at arbitrary spacetime points. The functional Taylor series is therefore more general than a local polynomial such as $b_1\delta + b_2\delta^2/2$. The familiar operator expansion is obtained only after the physical support and tensor structure of the kernels have been restricted. The normalization by $\bar{n}_g$ makes each kernel a fractional abundance response.

1.2.1 Causality, spatial locality, and time nonlocality

Galaxy formation has a finite spatial range and a finite duration. These two facts lead to different approximations. For wavelengths much longer than the formation scale R_*, the spatial dependence can be expanded in derivatives. The formation time, however, need not be short compared with a Hubble time, so there is no analogous reason to replace the entire temporal history by an instantaneous response.

The galaxy observed at $(\mathbf{x}, \eta)$ is built from matter elements whose trajectory is

$$\mathbf{x}_{\rm fl}(\eta'; \mathbf{x}, \eta) = \mathbf{x} - \int_{\eta'}^{\eta} \mathrm{d}\eta'' \, \mathbf{v}(\mathbf{x}_{\rm fl}(\eta''), \eta''). \tag{1.11}$$

The spacetime picture of Sec. 2.5 of[1] is local over a spatial scale R_* but generally nonlocal over a formation time of order the Hubble time. At leading order in the spatial derivative expansion, the support of the response kernel lies on the past fluid trajectory:

$$\begin{aligned}\mathcal{R}^{(N)}_{A_1 \cdots A_N}(x; x_1, \dots, x_N) \simeq {} & \prod_{a=1}^{N} \left[\Theta(\eta - \eta_a) \delta_D^{(3)}(\mathbf{x}_a - \mathbf{x}_{\rm fl}(\eta_a; \mathbf{x}, \eta)) \right] \\ & \times b_{A_1 \cdots A_N}(\eta; \eta_1, \dots, \eta_N) + \mathcal{O}(R_*^2 \nabla^2).\end{aligned} \tag{1.12}$$

The spatial Dirac distributions in Eq. 1.12 are the leading term of the long-wavelength expansion; they do not state that the microscopic formation process occurs at a mathematical point. They state that, after coarse graining, all long fields are evaluated along the worldline of the same matter element. The remaining kernel $b_{A_1 \cdots A_N}(\eta; \eta_1, \dots, \eta_N)$ retains the memory of when those fields acted. Substitution into Eq. 1.10 yields the explicitly time-nonlocal form

$$\begin{aligned}\delta_g^{\rm det}(\mathbf{x}, \eta) = {} & \sum_{N=1}^{\infty} \frac{1}{N!} \int^{\eta} \mathrm{d}\eta_1 \cdots \mathrm{d}\eta_N \, b_{A_1 \cdots A_N}(\eta; \eta_1, \dots, \eta_N) \\ & \times \prod_{a=1}^{N} \mathcal{O}_L^{A_a}(\mathbf{x}_{\rm fl}(\eta_a), \eta_a) + \mathcal{O}(R_*^2 \nabla^2).\end{aligned} \tag{1.13}$$

This equation is local in the sense relevant to the equivalence principle: the long fields are evaluated in the neighborhood of the same freely falling fluid element. It is not local in time.

Evaluating the operators on the fluid trajectory is also required by advection. A galaxy observed at $\mathbf{x}$ did not in general occupy the same comoving coordinate at earlier times. Replacing $\mathbf{x}_{\mathrm{fl}}(\eta_a)$ by the final Eulerian position would omit the transport of the long-wavelength environment and would generate an incomplete operator basis at nonlinear order.

The first spatial correction is generated by expanding the response around the trajectory. Isotropy eliminates a one-derivative scalar, so the leading scalar correction has the form

$$\delta_g \supset b_{\nabla^2\delta}\frac{\nabla^2\delta}{k_*^2}, \qquad k_* \sim R_*^{-1}. \tag{1.14}$$

In Fourier space this gives an analytic correction proportional to k^2/k_*^2.

Higher-derivative bias is therefore controlled by the ratio of the observed wavelength to the physical size of the formation region. It is distinct from loop corrections generated by nonlinear evolution: both can scale as powers of k, but they represent different short-distance information and carry independent coefficients.

1.2.2 The homogeneous response and the usual bias parameters

A spatially homogeneous long density perturbation Δ is the simplest background. The conditional mean abundance is then an ordinary function $\bar{n}_g(\Delta, \eta)$, and Eq. 1.9 reduces to

$$\boxed{b_N(\eta) \equiv \frac{1}{\bar{n}_g(0,\eta)} \left.\frac{\partial^N \bar{n}_g(\Delta,\eta)}{\partial \Delta^N}\right|_{\Delta=0}} \tag{1.15}$$

with

$$\frac{\bar{n}_g(\Delta)}{\bar{n}_g(0)} - 1 = b_1\Delta + \frac{b_2}{2}\Delta^2 + \cdots. \tag{1.16}$$

This is the separate-universe response definition for isotropic density bias. A tidal bias coefficient is obtained instead from a homogeneous *anisotropic* background and cannot be inferred from derivatives with respect to the scalar Δ alone.

The separate-universe language is useful because a sufficiently long isotropic mode changes the local expansion history rather than producing an appreciable spatial gradient across the formation region. Measuring how the abundance changes in that modified background gives the same large-scale coefficient that appears in correlation functions, provided the response is defined at fixed local proper time and with the same physical sample selection.

The derivative in Eq. 1.15 is taken at fixed physical definition of the sample and at fixed local proper time. Holding coordinate time, observed redshift, or global scale factor fixed gives a different derivative. This distinction is hidden in Newtonian notation and will become explicit in Lectures 2–3.

1.2.3 Local gravitational operators

In a freely falling frame a constant potential and a uniform acceleration can be removed. That statement is the Newtonian shadow of the equivalence principle, and Chapter 4 will derive its exact form by constructing coordinates in which the connection vanishes along a whole worldline. What survives the removal is the second derivative of the potential, so to leading order in derivatives the local gravitational data are contained in the Hessian,

$$\partial_i \partial_j \Phi = \frac{1}{3}\delta_{ij}\nabla^2\Phi + \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\Phi. \tag{1.17}$$

Using Poisson's equation, its trace is the density. Define the dimensionless trace-free tidal field

$$K_{ij} \equiv \left(\frac{\partial_i\partial_j}{\nabla^2} - \frac{1}{3}\delta_{ij}\right)\delta, \qquad K^i{}_i = 0. \tag{1.18}$$

A scalar tracer abundance cannot respond linearly to K_{ij} in an isotropic ensemble. The first scalar tidal invariant is $K_{ij}K^{ij}$. A shape tensor can respond linearly to K_{ij}; this is the response-theory origin of intrinsic alignment developed in Chapter 5.

The restriction is a statement about representations of the rotation group. A scalar abundance has no free spatial index with which to contract one trace-free rank-two tensor, whereas a shape observable already transforms as such a tensor. The same long-wavelength tide can therefore enter scalar clustering only at quadratic order but enter intrinsic shape at linear order.

Perturbative time evolution allows the time integrals in Eq. 1.13 to be reorganized into a basis of operators evaluated at the final time. Through second order and at leading derivative order,

$$\delta_g = b_1\delta + \frac{b_2}{2}[\delta^2] + b_{K^2}[K_{ij}K^{ij}] + b_{\nabla^2\delta}\frac{\nabla^2\delta}{k_*^2} + \epsilon + \cdots. \tag{1.19}$$

The brackets denote renormalized composite operators. Operationally, the renormalized coefficients are the large-scale responses defined above; they do not depend on an arbitrary smoothing scale used to construct the bare operators.

For example, the bare field δ^2 contains a nonzero mean and short-mode contributions that can mimic lower-order operators. Renormalization subtracts those pieces so that a coefficient such as b_2 measures a response to a long background rather than an artifact of how the density was smoothed. This is why the response definition and the operator basis should be specified together.

At tree level and for $k \ll k_*$,

$$\begin{aligned} P_{gm}(k) &= b_1 P_{mm}(k) + \mathcal{O}(k^2 P_{mm}), \\ P_{gg}(k) &= b_1^2 P_{mm}(k) + P_{\epsilon\epsilon}^{\{0\}} + \mathcal{O}(k^2 P_{mm}, k^2 P_{\epsilon\epsilon}). \end{aligned} \tag{1.20}$$

The response coefficient b_1 is a property of the selected population, not a prediction of gravity alone.

The constant term $P_{\epsilon\epsilon}^{\{0\}}$ is the leading large-scale stochastic contribution. It need not equal the Poisson value $1/\bar{n}_g$: exclusion, satellite occupation, conservation laws, and sample construction can change it. The tracer–matter cross spectrum does not contain this leading white auto-power; under the usual large-scale constraints on stochastic cross-correlations, it isolates the linear response b_1 more directly than the tracer auto-spectrum.

The functional derivative specifies what a bias parameter means. The local operator expansion specifies how the response can appear in long-wavelength correlation functions. The two statements are related but not interchangeable: a list of allowed operators without a response definition leaves the coefficients scheme-dependent, while a response to only a homogeneous density does not determine tidal or higher-derivative bias.

1.3 The real-to-redshift-space map

A spectroscopic catalogue labels a source by its observed redshift rather than by its real-space radial coordinate. In the Newtonian approximation the peculiar velocity is the only perturbation to this label. Converting the Doppler shift into a radial distance with the background relation produces a coordinate map; number conservation then supplies its Jacobian.

> **Historical note: Kaiser 1987**
>
> The linear result derived in this section is due to[6], who pointed out that the coherent infall of matter towards overdense regions compresses structures along the line of sight and therefore amplifies the measured clustering in a way that depends on orientation. The observation turned what had looked like a nuisance into a measurement: the amplitude of the anisotropy is set by the growth rate f, so redshift-space distortion became one of the standard probes of the growth of structure. The relativistic completion in Chapter 3 shows that the Kaiser term is one entry in a longer list, and identifies which of the remaining terms are suppressed on subhorizon scales.

Adopt the distant-observer approximation with a fixed line of sight $\hat{\mathbf{n}}$. To leading Newtonian order, the Doppler perturbation to redshift is interpreted through the background distance-redshift relation as the map

$$\tilde{\mathbf{x}} = \mathbf{x} + u_\parallel(\mathbf{x})\hat{\mathbf{n}}, \qquad u_\parallel \equiv \frac{v_\parallel}{\mathcal{H}}, \qquad v_\parallel \equiv \hat{\mathbf{n}}\cdot\mathbf{v}. \tag{1.21}$$

The Jacobian matrix is

$$\frac{\partial \tilde{x}^i}{\partial x^j} = \delta^i{}_j + \hat{n}^i \partial_j u_\parallel. \tag{1.22}$$

The perturbation to the identity is a rank-one matrix, the outer product of the fixed line of sight with the gradient of the radial displacement. This is a direct consequence of the plane-parallel approximation: only the radial coordinate is relabelled, so only one row of the deformation is populated. The determinant of such a matrix is elementary. The matrix-determinant lemma, $\det(I + \mathbf{a}\mathbf{b}^{\mathsf{T}}) = 1 + \mathbf{b}\cdot\mathbf{a}$, gives

$$J_s \equiv \det\left(\frac{\partial \tilde{x}^i}{\partial x^j}\right) = 1 + \partial_\parallel u_\parallel = 1 + \frac{1}{\mathcal{H}}\partial_\parallel v_\parallel. \tag{1.23}$$

Number conservation in a one-to-one region of the map implies

$$[1 + \delta_g^s(\tilde{\mathbf{x}})]\, \mathrm{d}^3\tilde{x} = [1 + \delta_g(\mathbf{x})]\, \mathrm{d}^3 x, \tag{1.24}$$

and therefore

$$1 + \delta_g^s(\tilde{\mathbf{x}}) = \frac{1 + \delta_g(\mathbf{x})}{1 + \mathcal{H}^{-1}\partial_\parallel v_\parallel(\mathbf{x})}. \tag{1.25}$$

Even in the plane-parallel approximation, Eq. 1.25 contains two nonlinear effects: the determinant in the denominator and the fact that the real-space fields must be evaluated at the inverse-mapped position $\mathbf{x}(\tilde{\mathbf{x}})$. Linear theory keeps only the first variation of each factor and may set the two positions equal inside a first-order field. The argument shift $\mathbf{x}(\tilde{\mathbf{x}})$ is essential beyond linear order. At linear order,

$$\delta_g^s = \delta_g - \frac{1}{\mathcal{H}}\partial_\parallel v_\parallel. \tag{1.26}$$

Using Eq. 1.3,

$$\partial_\parallel v_\parallel(\mathbf{k}) = -\mathcal{H} f \mu^2 \delta(\mathbf{k}), \qquad \mu \equiv \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}. \tag{1.27}$$

At leading bias order,

$$\boxed{\delta_g^s(\mathbf{k}) = (b_1 + f\mu^2)\delta(\mathbf{k})} \tag{1.28}$$

and

$$P_{gg}^s(k, \mu) = \left(b_1 + f\mu^2\right)^2 P_L(k). \tag{1.29}$$

The μ dependence is not an intrinsic anisotropy of the underlying matter statistics. It is generated by using the observer's line of sight to define the radial coordinate. Modes parallel to the line of sight receive the full velocity-gradient contribution; transverse modes do not. It is conventional to compress this angular dependence into Legendre multipoles,

$$P_\ell(k) = \frac{2\ell + 1}{2} \int_{-1}^{1} \mathrm{d}\mu \, P_{gg}^s(k, \mu) \, \mathcal{L}_\ell(\mu), \tag{1.30}$$

where $\mathcal{L}_\ell$ is a Legendre polynomial. At linear order only $\ell = 0, 2, 4$ are nonzero: The even multipoles are

$$\begin{aligned}
P_0 &= \left(b_1^2 + \frac{2}{3}b_1 f + \frac{1}{5}f^2\right) P_L, \\
P_2 &= \left(\frac{4}{3}b_1 f + \frac{4}{7}f^2\right) P_L, \\
P_4 &= \frac{8}{35}f^2 P_L.
\end{aligned} \tag{1.31}$$

The monopole mixes the overall clustering amplitude with growth, while the quadrupole and hexadecapole isolate the line-of-sight velocity response. In practice nonlinear velocities, wide-angle geometry, and relativistic light-cone terms modify this simple pattern, but the Kaiser result identifies the dominant short-distance limit.

1.4 What the Newtonian formula has suppressed

The two terms in Eq. 1.26 already have different origins:

$$\delta_m \xrightarrow{\text{formation response}} \delta_g \xrightarrow{\text{radial coordinate Jacobian}} \delta_g^s. \tag{1.32}$$

The Newtonian calculation leaves implicit

- the time hypersurface on which the functional derivative defining b_1 is taken;
- that a real catalogue is a flux of galaxy worldlines through the observer's past light cone, not a density on one Euclidean time slice;
- gravitational redshift, transverse deflection, area distortion, and observer terms;
- the difference between constant coordinate time, constant observed redshift, and constant source proper time.

The next lecture makes the slicing problem explicit, derives the redshift displacement and the cosmic clock, and replaces the Euclidean number-conservation statement Eq. 1.24 by a covariant pullback of the galaxy current three-form. Chapter 3 then supplies the spatial ruler determinant and assembles the temporal and spatial pieces into the observed density.

The Newtonian formula is therefore not discarded. It will be recovered by taking the subhorizon, weak-field, plane-parallel limit of the covariant observable. The purpose of the relativistic treatment is to state precisely which source density and which volume Jacobian enter before that limit is taken.

Summary

Key ideas. Galaxy bias is the functional response of a selected tracer population to its long-wavelength environment, Eq. 1.9, not a polynomial fitted to the matter field. Two consequences follow. The allowed operators are fixed by symmetry, causality, and the equivalence principle rather than by convenience, which is why a scalar abundance can respond to $K_{ij}K^{ij}$ but not to K_{ij}. And the response is local in space over the formation scale R_* while remaining nonlocal in time over a Hubble time, so the operators must be evaluated along the past fluid trajectory. Separately, redshift-space anisotropy is not a property of the matter statistics; it is manufactured by using the observer's line of sight to define the radial coordinate.

Main results.

- The functional Taylor expansion Eq. 1.10 and its local, time-nonlocal reduction Eq. 1.13.
- The separate-universe definition of the density bias coefficients, Eq. 1.15, taken at fixed local proper time and fixed physical sample definition.
- The renormalized operator basis Eq. 1.19, and the tree-level spectra Eq. 1.20.
- The Kaiser field and power spectrum, Eq. 1.28 and Eq. 1.29, with multipoles Eq. 1.31.

What the next chapter builds on. The Newtonian derivation left four things unstated, listed above Eq. 1.32: the time hypersurface on which the response is defined, the fact that a catalogue is a flux through the past light cone rather than a density on a Euclidean slice, the omission of gravitational redshift and propagation effects, and the difference between coordinate time, observed redshift, and source proper time. Chapter 2 makes the slicing problem explicit, derives the redshift displacement and the cosmic clock, and replaces Eq. 1.24

by the pullback of a covariant current three-form. Chapter 3 then supplies the spatial ruler determinant and assembles the temporal and spatial pieces into the observed density.

2 Gauge freedom and covariant number counts

Where we are

Chapter 1 ended with a list of things the Newtonian derivation had left unstated. Three of the four concern time: which hypersurface the bias response is defined on, which hypersurface a catalogue bin corresponds to, and how the two differ. The fourth is structural. A catalogue is not a density on a Euclidean time slice; it is a set of worldlines crossing the observer's past light cone. The Newtonian statement $[1 + \delta_g^s(\tilde{\mathbf{x}})]\,\mathrm{d}^3\tilde{x} = [1 + \delta_g(\mathbf{x})]\,\mathrm{d}^3 x$ presumes both a preferred simultaneity and a nondegenerate three-volume. Neither exists on a null surface.

General relativity introduces coordinate freedom, but the observational problem is not solved by selecting a preferred gauge. The catalogue count can be defined exactly as a geometric flux through the observer's past light cone. Perturbation theory is then an expansion of this exact object. Gauge-dependent density, redshift, and displacement terms appear because the exact object has been split into background and perturbation, not because the measured count is ambiguous.

The chapter is arranged in two halves, in the order in which a working calculation meets the difficulties rather than in the order in which they are best explained.

The first half is perturbative and concrete. It defines a gauge transformation and shows immediately that a density perturbation has no meaning until a slicing is named. It then works out how the metric perturbations transform, how the observed redshift determines the displacement of the true source event from the fiducial one, and how the proper time elapsed at the source differs from the time the fiducial cosmology assigns to that redshift. The dimensionless field measuring the mismatch is the cosmic clock $\mathcal{T}$, and it is gauge invariant. By the end of this half the reader has a working formalism and one nagging observation: several manifestly gauge-dependent quantities have combined into invariant ones, and nothing so far explains why.

The second half explains why. The object that counts worldlines through any three-surface, null or spacelike, is a differential three-form, and integrating it requires no metric on the surface at all. Written that way the count could not have depended on coordinates, and the cancellations of the first half become forced rather than fortunate. Along the way the light-cone geometry is put in a frame adapted to the observation, which separates the radial direction along the ray from the two transverse screen directions and shows in advance which distortions can change a volume.

What this chapter does not supply is the spatial part of the volume distortion itself. That is the business of Chapter 3.

Figure placeholder. Illustration of the pullback. On the left, a bundle of galaxy worldlines piercing a three-surface embedded in spacetime; on the right, the same bundle represented on the observed-coordinate manifold $(\tilde{z}, \tilde{\theta}, \tilde{\phi})$, with the embedding X drawn as the arrow between them and the three tangent vectors $\partial X^\mu/\partial y^a$ marked on the surface.

2.1 Gauge transformations and the slicing of a density perturbation

A gauge transformation in cosmological perturbation theory is a relabelling of spacetime events,

$$x^\mu \mapsto x^\mu + \xi^\mu. \tag{2.1}$$

The transformation is passive. No field is moved and no physical configuration is altered; only the coordinates assigned to events change. Throughout these notes, a tilde is reserved for quantities inferred from the observed labels $(\tilde{z}, \hat{\mathbf{n}})$ using the background cosmology. Gauge transformations are therefore written with $\mapsto$ rather than by decorating the transformed field with another tilde.

In the scalar sector the generator is

$$\xi^\mu = (T, \partial^i L), \tag{2.2}$$

where T shifts the time slicing and L relabels positions within a slice. Almost all of the difficulty addressed in this chapter comes from the first of these.

The reason is that a perturbation is not a property of a single event. Splitting a field into a background and a perturbation requires a background, the background is a function of time alone, and reading off "the time" of an event requires a slicing. Change the slicing and the perturbation changes, even though nothing physical has moved. The density makes this concrete.

The word "density" can obscure the actual comparison. The scalar $\rho(x)$ at one event is invariant, but a perturbation compares it with a homogeneous reference value $\bar{\rho}$ evaluated at a second, background event. A time slicing supplies the rule that pairs these events. Changing the slicing changes the pair and hence changes the perturbation.

The value of a physical scalar at a specified event is unchanged by a coordinate relabelling. The split

$$\rho(\eta, \mathbf{x}) = \bar{\rho}(\eta) + \delta\rho(\eta, \mathbf{x}) \tag{2.3}$$

compares the physical density with the background density at the same coordinate time. Under the time shift generated by T, its perturbative representation changes as

$$\delta\rho \mapsto \delta\rho - \bar{\rho}' T. \tag{2.4}$$

For pressureless matter, $\bar{\rho}_m' = -3\mathcal{H}\bar{\rho}_m$, so

$$\boxed{\delta_m \mapsto \delta_m + 3\mathcal{H}T}. \tag{2.5}$$

For a tracer with mean physical number density $\bar{n}_g$,

$$\delta_g \mapsto \delta_g - \frac{\bar{n}_g'}{\bar{n}_g} T. \tag{2.6}$$

These formulas do not say that density is unphysical. They say that the perturbation requires a rule for identifying simultaneous events.

The required rule depends on the physical question. Galaxy formation is naturally compared at equal local proper time, whereas a catalogue bin is constructed at equal observed redshift. The difference between these two hypersurfaces is measurable and is precisely the cosmic-clock perturbation introduced below.

2.2 Metric perturbations and scalar gauge transformations

To connect the geometric argument with standard cosmological perturbation variables, we now write the most general linear scalar perturbation of a flat FLRW metric. The four scalar functions A, B, D, E are related by two scalar coordinate freedoms; the Einstein and matter equations further constrain the dynamical combinations. A choice of gauge imposes two coordinate conditions; it does not select a different spacetime.

Write the linearly perturbed flat FLRW metric as

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2A)\mathrm{d}\eta^2 - 2B_i \mathrm{d}\eta \mathrm{d}x^i + (\delta_{ij} + h_{ij})\mathrm{d}x^i \mathrm{d}x^j\right]. \tag{2.7}$$

In the scalar sector,

$$B_i = \partial_i B, \qquad h_{ij} = 2D\delta_{ij} + 2\partial_i \partial_j E. \tag{2.8}$$

Under the generator Eq. 2.2, linear metric perturbations obey

$$\delta g_{\mu\nu} \mapsto \delta g_{\mu\nu} - \mathcal{L}_\xi \bar{g}_{\mu\nu}. \tag{2.9}$$

Evaluating the Lie derivative gives

$$\begin{aligned} A &\mapsto A - \mathcal{H}T - T', \\ B &\mapsto B + L' - T, \\ D &\mapsto D - \mathcal{H}T, \\ E &\mapsto E - L. \end{aligned} \tag{2.10}$$

If $u^i = a^{-1}\partial^i v$ at linear order,

$$v \mapsto v + L'. \tag{2.11}$$

Table C.1 collects these transformations together with those of the displacement, clock, and density variables introduced later in the chapter, in a single passive convention. It is worth consulting before comparing any formula here with one taken from another source, because the sign of the generator and the sign of the spatial potential both vary across the literature.

Historical note: Bardeen 1980

The systematic construction of combinations of metric and matter perturbations that are unchanged under Eq. 2.1 is due to[7]. Before that work, disagreements between calculations performed in different gauges were a recurring source of confusion, and spurious growing modes in synchronous gauge were a familiar hazard. Bardeen's variables made it possible to state which parts of a perturbation carry physical information without first committing to coordinates. The construction in this chapter reaches the same conclusion from the opposite direction: rather than building invariants out of the fields, it defines the measurement covariantly and lets the invariance follow.

We use two common gauges for checks:

- synchronous-comoving gauge: $A = B = v = 0$;
- conformal-Newtonian gauge: $B = E = 0$, $A = \Psi$, and $D = \Phi$ in

$$\mathrm{d}s^2 = a^2 \left[-(1 + 2\Psi)\mathrm{d}\eta^2 + (1 + 2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] . \tag{2.12}$$

The spatial-potential sign in Eq. 2.12 is opposite to the convention $g_{ij} = a^2(1-2\Phi_{\text{common}})\delta_{ij}$.

Synchronous-comoving and conformal-Newtonian gauges are useful because different terms simplify in each. Agreement of the final observable between them is a check, but gauge invariance is already guaranteed by the pullback construction. One should not interpret a term that vanishes in one gauge as a physically absent mechanism unless the complete observable combination vanishes.

2.3 Ray tracing from the observer to the source

The gauge discussion above identifies the problem: the coordinate position of a source is not an observable. The operational solution is not to choose a preferred coordinate system, but to formulate the observation as an initial-value problem whose boundary data are fixed in the observer's local frame. The observer measures a photon energy and a direction, and the metric then determines which source event produced that photon.

Let the observation event be x_o and let the observer carry an orthonormal tetrad $e_{\hat{a}}{}^{\mu}$ with

$$e_{\hat{0}}{}^{\mu} = u_o^{\mu}, \qquad g_{\mu\nu}e_{\hat{a}}{}^{\mu}e_{\hat{b}}{}^{\nu} = \eta_{\hat{a}\hat{b}}. \tag{2.13}$$

The measured line of sight is the unit spatial vector

$$\hat{n}^{\mu} = \hat{n}^{\hat{i}}e_{\hat{i}}{}^{\mu}, \qquad u_o \cdot \hat{n} = 0, \qquad \hat{n} \cdot \hat{n} = 1, \tag{2.14}$$

where $\hat{\mathbf{n}}$ points from the observer toward the source. We take k^{μ} to be the future-directed photon wavevector, so at the observer

$$\boxed{k_o^{\mu} = E_o\left(u_o^{\mu} - \hat{n}^{\mu}\right)}, \qquad E_o \equiv -u_{o\mu}k_o^{\mu} > 0. \tag{2.15}$$

The overall normalization E_o only fixes the affine parameter and may be set to unity. The relative spatial components of k_o^{μ} are fixed by the measured direction. Thus the local observation supplies both the endpoint and all initial data needed for the null geodesic.

The photon trajectory is obtained from

$$\frac{\mathrm{d}x^{\mu}}{\mathrm{d}\lambda} = k^{\mu}, \qquad \frac{\mathrm{D}k^{\mu}}{\mathrm{d}\lambda} = 0, \tag{2.16}$$

or, in coordinates,

$$\frac{\mathrm{d}k^\mu}{\mathrm{d}\lambda} = -\Gamma^\mu_{\alpha\beta} k^\alpha k^\beta. \tag{2.17}$$

Starting from (x_o, k_o), one integrates this system backward along the ray. Equivalently, one keeps k^μ future directed and integrates toward decreasing affine parameter. The source affine parameter λ_s is not specified in advance. It is selected by the measured redshift through

$$1 + \tilde{z} = \frac{-u_{s\mu} k_s^\mu}{-u_{o\mu} k_o^\mu} = \frac{E_s}{E_o}. \tag{2.18}$$

The first point on the chosen source congruence for which Eq. 2.18 holds is the physical emission event,

$$x_s^\mu(\tilde{z}, \hat{\mathbf{n}}) = x^\mu(\lambda_s). \tag{2.19}$$

The procedure can be summarized as

$$\boxed{(x_o, u_o, E_o, \hat{\mathbf{n}}) \xrightarrow{\mathrm{D}k/\mathrm{d}\lambda=0} (x_s, k_s) \xrightarrow{E_s/E_o=1+\tilde{z}} x_s(\tilde{z}, \hat{\mathbf{n}})}. \tag{2.20}$$

This is the operational content behind the displacement field used below. The background FLRW model assigns a fiducial event $\tilde{x}^\mu(\tilde{z}, \hat{\mathbf{n}})$ to the same observed labels, while the geodesic integration gives the actual event. Their coordinate difference is $\Delta x^\mu = x_s^\mu - \tilde{x}^\mu$. Under a gauge transformation both coordinate descriptions change, but the observer data, the null ray, and the physical source event do not. The transformation of Δx^μ is therefore forced to compensate the transformation of the fields evaluated at the fiducial event.

At linear order the exact boundary-value problem is expanded about the background ray. The perturbation to the photon energy gives $\Delta \ln a$, while the integrated perturbations to the trajectory give $\Delta x_\parallel$ and $\Delta x_\perp^i$. The formulas in the next section are therefore not independent projection corrections. They are different components of one ray-tracing solution with boundary conditions fixed at the observer.

! The observer sets the boundary conditions

A cosmological observable is not obtained by choosing a source coordinate and propagating a photon toward the observer. The measured energy and direction are fixed at the observer, the null geodesic is integrated backward, and the observed redshift selects the emission event. This ordering is what ensures that source, propagation, and observer terms belong to one prediction.

2.4 Observed redshift and the displacement field

The background map assigns a spacetime event to each observed pair $(\tilde{z}, \hat{\mathbf{n}})$. Perturbations change both the emission time and the spatial position of the actual source relative to that fiducial event. The displacement field Δx^μ records this difference. Its time component is conveniently expressed as a perturbation to the emission scale factor.

The observed redshift and direction define the fiducial event

$$\tilde{x}^\mu = (\eta_0 - \tilde{\chi}, \tilde{\chi}\hat{\mathbf{n}}), \qquad \tilde{a} = (1+\tilde{z})^{-1}. \tag{2.21}$$

The physical source event is

$$x^\mu = \tilde{x}^\mu + \Delta x^\mu. \tag{2.22}$$

The observed labels $(\tilde{z}, \hat{\mathbf{n}})$ are measured quantities and are therefore unchanged by a relabelling of coordinates, so the fiducial event $\tilde{x}^\mu$ carries no gauge dependence. All of it is absorbed by the displacement, which transforms as

$$\Delta x^\mu \mapsto \Delta x^\mu + \xi^\mu. \tag{2.23}$$

This sign, opposite to the one carried by field perturbations at a fixed coordinate point, is what makes the observed count invariant; Sec. 2.11 shows that the two are the two halves of a single geometric statement.

Define

$$\Delta \ln a \equiv \ln \frac{a(x^0)}{\tilde{a}}. \tag{2.24}$$

Thus $\Delta \ln a$ is evaluated at fixed measured redshift; it is not an independent fluctuation of the observed redshift. It answers the question: by how much does the background scale factor at the actual emission event differ from the scale factor that the fiducial FLRW model assigns to the measured redshift? Since the observed labels are held fixed while Δx^0 transforms as in Eq. 2.23,

$$\Delta \ln a \mapsto \Delta \ln a + \mathcal{H}T. \tag{2.25}$$

Operationally, redshift is the photon-energy ratio

$$1+\tilde{z} = \frac{(u_\mu k^\mu)_s}{(u_\mu k^\mu)_o}. \tag{2.26}$$

Linearizing the source and observer energies and integrating the perturbed null geodesic gives, along the unperturbed ray,

$$\begin{aligned} \Delta \ln a = A_o - A_s + v_{\|s} - v_{\|o} + \int_0^{\tilde{\chi}} \mathrm{d}\chi \left[-A' + \frac{1}{2}h'_{\|} + B'_{\|}\right] \\ - H_0 \int_0^{t_o} \mathrm{d}t\, A(\mathbf{0}, \bar{\eta}(t)). \end{aligned} \tag{2.27}$$

The last term converts fixed observer coordinate time into fixed observer proper time. In conformal-Newtonian gauge,

$$
\begin{aligned}
(\Delta \ln a)_{\mathrm{cN}} = \Psi_o - \Psi_s + v_{\parallel s} - v_{\parallel o} + \int_0^{\widetilde{\chi}} \mathrm{d}\chi \, (\Phi' - \Psi') \\
- H_0 \int_0^{t_o} \mathrm{d}t \, \Psi(\mathbf{0}, \bar{\eta}(t)).
\end{aligned}
\tag{2.28}
$$

The individual endpoint, line-of-sight, and observer contributions are gauge-dependent. Their combination is fixed by Eq. 2.26.

The endpoint terms describe the gravitational and Doppler shifts at source and observer. The line-of-sight integral describes the change in photon energy as the metric evolves along the ray. The final observer term enforces that the observation is made at fixed observer proper time rather than at fixed coordinate time. Dropping observer terms can leave spurious monopole or dipole contributions and can spoil gauge invariance even when the source terms are otherwise correct.

2.5 Source proper time at fixed observed redshift: the cosmic clock

The bias expansion of Chapter 1 compares local tracer populations at equal proper time. A catalogue, by contrast, compares sources at equal measured redshift. The quantity required for the bias relation is therefore not the coordinate time of the source, but the proper time accumulated along its worldline relative to the background event assigned to the same observed redshift.

Let Σ_{ini} be a physically specified early hypersurface from which this proper time is measured. We set the zero of t_F on Σ_{ini}. Along a nonrelativistic source worldline $\mathbf{x}_{\mathrm{fl}}(\eta)$,

$$
\mathrm{d}t_F = a(1 + A)\mathrm{d}\eta \tag{2.29}
$$

to first order; terms involving the peculiar velocity enter only quadratically. Thus

$$
t_F(\eta, \mathbf{x}) = \bar{t}(\eta) + I_A(\eta, \mathbf{x}), \tag{2.30}
$$

where

$$
\begin{aligned}
\bar{t}(\eta) &\equiv \int_{\eta_{\mathrm{ini}}}^{\eta} \mathrm{d}\eta' \, a(\eta'), \\
I_A(\eta, \mathbf{x}) &\equiv \int_{\eta_{\mathrm{ini}}}^{\eta} \mathrm{d}\eta' \, a(\eta') A(\mathbf{x}_{\mathrm{fl}}(\eta'), \eta').
\end{aligned}
\tag{2.31}
$$

At linear order, replacing $\mathbf{x}_{\mathrm{fl}}(\eta')$ by the unperturbed spatial position inside the already first-order field A changes the integral only at second order.

The lower boundary needs one explicit convention. We use Σ_{ini} as the common physical reference surface and anchor the time relabelling there by requiring

$$
(aT)_{\mathrm{ini}} = 0. \tag{2.32}
$$

Equivalently, the two coordinate descriptions use the same zero of t_F on Σ_{ini}. This is a boundary convention for the gauge generator, not a dynamical assumption. With an arbitrary lower-boundary convention, the transformation of I_A retains an explicit $(aT)_{\rm ini}$ term. Indeed, from Eq. 2.10,

$$aA \mapsto aA - a\mathcal{H}T - aT' = aA - (aT)', \tag{2.33}$$

and hence

$$I_A \mapsto I_A - aT + (aT)_{\rm ini} = I_A - aT \qquad ((aT)_{\rm ini} = 0). \tag{2.34}$$

We can now define the clock operationally. At fixed observed labels $(\tilde{z}, \hat{\mathbf{n}})$, the background cosmology assigns the fiducial event $\tilde{x}^\mu$ and the background proper time $\bar{t}(\tilde{a})$. The actual source event is $x_s^\mu = \tilde{x}^\mu + \Delta x^\mu$. Their proper-time difference is

$$\Delta t_F(\tilde{z}, \hat{\mathbf{n}}) \equiv t_F(x_s) - \bar{t}(\tilde{a}). \tag{2.35}$$

Expanding about the fiducial event gives

$$\begin{aligned} \Delta t_F &= I_A(\tilde{x}) + a\Delta x^0 \\ &= I_A + \frac{\Delta \ln a}{H}, \end{aligned} \tag{2.36}$$

where the second line uses $\Delta \ln a = \mathcal{H}\Delta x^0$ and $\mathcal{H} = aH$. All background quantities in these equations are evaluated at $\tilde{a} = (1+\tilde{z})^{-1}$. The dimensionless cosmic clock of[3] is therefore

$$\boxed{\mathcal{T}(\tilde{z}, \hat{\mathbf{n}}) \equiv H\Delta t_F = HI_A + \Delta \ln a}. \tag{2.37}$$

The meaning is direct: $\mathcal{T}/H$ is the perturbation in the local cosmic age of the source event selected at fixed observed redshift. It is not an additional force or propagation effect. It records that the observed-redshift hypersurface and the equal-proper-time hypersurface intersect the source congruence at different events. This is why any locally defined population whose mean abundance evolves with proper time receives a clock contribution.

The gauge invariance is now transparent. Using Eq. 2.34 and Eq. 2.25,

$$\mathcal{T} \mapsto H(I_A - aT) + \Delta \ln a + \mathcal{H}T = \mathcal{T}. \tag{2.38}$$

2.6 Constant-redshift and constant-proper-time slicings

The two slicings are useful bookkeeping choices for the same physical configuration. In constant-observed-redshift gauge the source event has no time displacement relative to the background redshift label, so the entire mismatch appears as a proper-time perturbation. In constant-proper-time gauge the local source clock is unperturbed, so the same mismatch appears in $\Delta \ln a$. Neither gauge is preferred by the observable.

A time shift

$$T_z = -\frac{\Delta \ln a}{\mathcal{H}} \tag{2.39}$$

sends $\Delta \ln a \mapsto 0$. Denoting quantities in this slicing by a subscript z,

$$(\Delta \ln a)_z = 0, \qquad H(I_A)_z = \mathcal{T}.$$

Coordinate time then labels the constant-observed-redshift hypersurface, while the entire mismatch appears as a source proper-time perturbation.

A time shift

$$T_{\rm pt} = \frac{I_A}{a} \tag{2.40}$$

sends $I_A \mapsto 0$. In the constant-proper-time slicing,

$$(I_A)_{\rm pt} = 0, \qquad (\Delta \ln a)_{\rm pt} = \mathcal{T}. \tag{2.41}$$

Coordinate time then agrees with source proper time, while the entire mismatch appears in the relation between emission time and observed redshift.

For conserved matter,

$$\begin{aligned} \delta_m^{\rm or} &\equiv \delta_m - 3\Delta \ln a, \\ \delta_m^{\rm pt} &\equiv \delta_m + 3HI_A, \end{aligned} \tag{2.42}$$

are separately gauge-invariant and satisfy

$$\boxed{\delta_m^{\rm or} = \delta_m^{\rm pt} - 3\mathcal{T}}. \tag{2.43}$$

The bias response of Chapter 1 is naturally defined at fixed source proper time, so $\delta_m^{\rm pt}$ is the density entering the relativistic bias relation. The catalogue, however, is binned at fixed observed redshift. Chapter 3 combines the clock with spatial ruler distortions to connect the two.

Eq. 2.43 isolates the mechanism behind the density gauge issue. The difference between the two gauge-invariant density perturbations is itself a gauge-invariant clock.

This completes the perturbative half of the chapter, and it is worth pausing on what has happened. Individually, δ_m, $\Delta \ln a$, and I_A are all gauge dependent. Yet $\mathcal{T}$, $\delta_m^{\rm or}$, and $\delta_m^{\rm pt}$ are not, and the invariance was verified in each case by substituting the transformation rules and watching terms cancel. Nothing in that procedure explains why the cancellations occur, and a calculation of the full number count involves many more terms of the same kind. The remainder of the chapter shows that the cancellations are not coincidences to be checked one at a time. They follow from the fact that the count was a single geometric object before it was split.

2.7 Galaxy number as the flux of a current three-form

A galaxy catalogue counts worldlines crossing a three-dimensional surface. The natural covariant object is therefore not a scalar density alone but a three-form that can be integrated over that surface. Contracting the number current with the spacetime volume form supplies exactly such an object and automatically includes the relative orientation between the current and the surface.

Let M be an oriented Lorentzian spacetime with volume form

$$\boldsymbol{\varepsilon} \equiv \sqrt{-g}\,\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3, \qquad \epsilon_{0123} = +1. \tag{2.44}$$

The galaxy current is

$$j = j^\mu \partial_\mu, \qquad j^\mu = n_g u_g^\mu, \qquad u_g^\mu u_{g\mu} = -1, \tag{2.45}$$

where n_g is the physical number density measured in the galaxy rest frame. Contracting the four-volume form with the current gives the current three-form

$$\boxed{\mathcal{J} \equiv \iota_j \boldsymbol{\varepsilon} = \star j^\flat} \tag{2.46}$$

with coordinate expression

$$\mathcal{J} = \frac{\sqrt{-g}}{3!}\,\epsilon_{\mu\nu\rho\sigma} j^\mu \mathrm{d}x^\nu \wedge \mathrm{d}x^\rho \wedge \mathrm{d}x^\sigma. \tag{2.47}$$

This is the covariant replacement for $(1+\delta_g)\mathrm{d}^3x$ in the Newtonian number-conservation argument.

At a given event, $\mathcal{J}$ contains both the rest-frame number density and the velocity with which the worldlines pierce the chosen hypersurface. These pieces should not be separated before the hypersurface is specified. On a surface orthogonal to u_g^μ the velocity factor reduces to unity; on a tilted or null surface it is part of the measured flux.

Cartan's identity gives its exterior derivative:

$$\begin{aligned} \mathrm{d}\mathcal{J} &= \mathrm{d}(\iota_j \boldsymbol{\varepsilon}) = \mathcal{L}_j \boldsymbol{\varepsilon} - \iota_j \mathrm{d}\boldsymbol{\varepsilon} \\ &= \left(\nabla_\mu j^\mu\right) \boldsymbol{\varepsilon}. \end{aligned} \tag{2.48}$$

Thus $\nabla_\mu j^\mu = 0$ is equivalent to $\mathrm{d}\mathcal{J} = 0$. If two hypersurfaces bound a four-volume without galaxy creation, destruction, or flux through the side boundary, Stokes' theorem gives the same total count on both. The functorial identities used here, and Cartan's formula in particular, are collected in Appendix B; every step in this section is one of those identities, with no input from the field equations.

For an evolving tracer population one may instead have $\nabla_\mu j^\mu \neq 0$ because objects enter or leave the selected sample. The three-form still defines the count on any one hypersurface; nonconservation then describes physical source evolution between hypersurfaces. The mean effect of this evolution is later encoded by b_e rather than being folded into the geometric volume distortion.

2.8 Pullback to a three-dimensional hypersurface

The embedding X specifies which spacetime event corresponds to each coordinate point y^a on the three-manifold Σ. The pullback X^* converts a spacetime form into a form that can be integrated using the coordinates y^a. No coordinate system on spacetime is preferred by this operation; changing spacetime coordinates changes the components of both the form and the embedding derivatives while leaving the pulled-back form unchanged.

Let Σ be a spacelike or null three-dimensional hypersurface, represented by an embedding

$$\mathsf{X} : \Sigma \longrightarrow M, \qquad y^a \longmapsto X^\mu(y^1, y^2, y^3). \tag{2.49}$$

The number of galaxy worldlines crossing Σ is

$$\boxed{N(\Sigma) = \int_\Sigma \mathsf{X}^* \mathcal{J}}. \tag{2.50}$$

This formula is meaningful for a null hypersurface even though the induced metric on a null hypersurface is degenerate and does not define a Riemannian three-volume.

This point is central for cosmology. A constant-time slice has a natural proper three-volume, but the past light cone does not. Trying to define the count as "density times the induced light-cone volume" therefore introduces an unnecessary difficulty. The current three-form already has the correct degree and can be pulled back directly.

The pullback acts on the coordinate one-forms as

$$\mathsf{X}^*(\mathrm{d}x^\mu) = \mathrm{d}(X^\mu(y)) = \frac{\partial X^\mu}{\partial y^a}\mathrm{d}y^a. \tag{2.51}$$

Substituting into Eq. 2.47 and collecting the coefficient of $\mathrm{d}y^1 \wedge \mathrm{d}y^2 \wedge \mathrm{d}y^3$ gives

$$\boxed{\mathsf{X}^* \mathcal{J} = \sqrt{-g}\,\epsilon_{\mu\nu\rho\sigma} j^\mu \frac{\partial X^\nu}{\partial y^1}\frac{\partial X^\rho}{\partial y^2}\frac{\partial X^\sigma}{\partial y^3}\mathrm{d}y^1 \wedge \mathrm{d}y^2 \wedge \mathrm{d}y^3} \tag{2.52}$$

for the orientation chosen in Eq. 2.44. Reversing the orientation of Σ reverses the sign of the form; the physical count uses the orientation for which the integrand is positive.

The antisymmetric contraction in Eq. 2.52 is the covariant analogue of a Jacobian determinant. The three tangent vectors $\partial X^\mu/\partial y^a$ span the hypersurface, while j^μ supplies the fourth direction. Their oriented four-volume measures how many worldlines cross the coordinate cell $\mathrm{d}^3 y$.

2.8.1 Spacelike check

Suppose Σ is spacelike with future-directed unit normal n^μ and induced volume element $\mathrm{d}\Sigma$. The pullback reduces to

$$\mathsf{X}^* \mathcal{J} = -(j \cdot n)\mathrm{d}\Sigma. \tag{2.53}$$

On a hypersurface orthogonal to the galaxy four-velocity, $n^\mu = u_g^\mu$, and hence

$$-(j \cdot u_g)\mathrm{d}\Sigma = n_g \mathrm{d}\Sigma. \tag{2.54}$$

This recovers the rest-frame number density. On the past light cone there is no unit normal, but Eq. 2.52 remains valid without modification.

The spacelike check also shows what changes when the surface is not orthogonal to the galaxy flow: the factor $-(j \cdot n)$ contains a Lorentz projection between the worldlines and the surface normal. The null case is the continuous limit in which a unit normal ceases to exist, while the differential-form expression remains regular.

2.9 Radial and tangential structure on the light cone

Eq. 2.52 is correct but opaque. A single antisymmetric contraction of four indices hides the fact that the light cone has a preferred direction, namely the one along which the photons travel, and that the remaining two directions are the observer's screen. Adapting the frame to that split, following[4], makes the structure of the count visible before any perturbation is introduced, and it explains in advance why only two of the six ruler distortions of Chapter 3 can change a volume.

Let k^μ be tangent to the null generators of Σ_{τ_o}, that is, to the photon trajectories reaching the observer. Complete it to a null tetrad by choosing a second null vector l^μ and a complex vector m^μ spanning the two directions transverse to the ray:

$$k \cdot k = l \cdot l = m \cdot m = 0, \qquad k \cdot l = -1, \qquad m \cdot \bar{m} = 1, \tag{2.55}$$

with $k \cdot m = l \cdot m = 0$. The complex vector is a repackaging of a real orthonormal screen basis,

$$\begin{aligned} m^\mu &= \frac{1}{\sqrt{2}} \left(\hat{e}_1^\mu + \mathrm{i}\, \hat{e}_2^\mu \right), \\ \mathrm{d}\zeta &= \frac{1}{\sqrt{2}} \left(\theta^1 + \mathrm{i}\, \theta^2 \right). \end{aligned} \tag{2.56}$$

Here $\hat{e}_1, \hat{e}_2$ are the real orthonormal screen vectors and θ^1, θ^2 their dual one-forms. They are the same transverse pair used to write the deformation matrix in Sec. 3.5.

The tangent space of the light cone at a point is spanned by k, m, and $\bar{m}$. A null surface is degenerate precisely because k is simultaneously tangent and normal, which is what obstructed defining an induced three-volume in Sec. 2.8. The current three-form does not care. Evaluating it on the three tangent directions and expanding the current in the tetrad basis,

$$j^\mu = -(j \cdot l) k^\mu - (j \cdot k) l^\mu + (j \cdot \bar{m}) m^\mu + (j \cdot m) \bar{m}^\mu, \tag{2.57}$$

every term except the one proportional to l^μ repeats a slot already occupied and is annihilated by the antisymmetry. The count therefore sees the current only through the single scalar

$$n_{g,\mathrm{lc}} \equiv k \cdot j = n_g \, (k \cdot u_g), \tag{2.58}$$

which is the density coefficient associated with the chosen affine parametrization of the light cone. By itself it rescales when the affine parameter is rescaled; the invariant object is the product with the affine measure. Parametrizing the null generators by an affine parameter λ

with $k^\mu = \mathrm{d}x^\mu/\mathrm{d}\lambda$, the combination $-(k \cdot u_g)\mathrm{d}\lambda$ is exactly the proper length increment along the beam measured in the galaxy rest frame,

$$\mathrm{d}\ell \equiv -k^\mu u_{g\mu}\, \mathrm{d}\lambda > 0, \tag{2.59}$$

so the factor $k \cdot u_g$ in Eq. 2.58 is not an additional weight. It is what converts an affine parameter into a physical length. The transverse factor is the screen area form

$$\boldsymbol{\sigma} \equiv \mathrm{i}\, \mathrm{d}\zeta \wedge \mathrm{d}\bar{\zeta} = \theta^1 \wedge \theta^2. \tag{2.60}$$

Using Eq. 2.56, the equality is a two-line wedge-product check: the complex packaging is bookkeeping, and $\boldsymbol{\sigma}$ is the ordinary proper area element of the screen. Combining it with the radial proper-length form gives

$$\boxed{\mathsf{X}^* \mathcal{J} = n_g\, \mathrm{d}\ell \wedge \boldsymbol{\sigma}}. \tag{2.61}$$

Equivalently, on the observer's past light cone,

$$\boxed{N(\Sigma_{\tau_o}) = \int_{\Sigma_{\tau_o}} n_g\, \mathrm{d}\ell \wedge \boldsymbol{\sigma}}. \tag{2.62}$$

As with Eq. 2.52, the orientation is fixed so that the integrand is positive.

Eq. 2.61 is the result. The count is a rest-frame number density multiplied by a radial proper length and a transverse proper area, with no metric on the null surface required anywhere.

Three things follow, and each is worth stating before the perturbative machinery starts.

The first is that only one component of the current survives. Of the four terms in Eq. 2.57 the pullback keeps only $k \cdot j$. The factor $-(k \cdot u_g)$ is the photon energy measured in the galaxy rest frame, but after Eq. 2.59 it belongs to the radial proper-length element rather than supplying a second weight. The same contraction also appears in the redshift ratio Eq. 2.26, so Doppler physics and radial number counting share one geometric ingredient.

The second is that the measure factorizes. The physical volume through which the worldlines pass is a radial proper length multiplied by a transverse proper area, and these are separately meaningful because the ray direction is singled out by the observation itself. This is the geometric origin of the split into longitudinal and transverse ruler distortions in Chapter 3.

The third is a prediction about which distortions can matter. A perturbation deforms the triad $(\hat{e}_\parallel, \hat{e}_1, \hat{e}_2)$, and the product in Eq. 2.61 responds to the determinant of that deformation. A radial stretch changes $\mathrm{d}\ell$ and a transverse stretch changes $\boldsymbol{\sigma}$, so both contribute at first order. Tilting the radial leg against the screen changes neither factor at first order, because a determinant is insensitive to off-diagonal entries until second order. This is exactly the statement, derived by other means in Eq. 3.28, that $\mathcal{B}_i$ is absent from the linear volume Jacobian.

Why the null tetrad is worth the notation

The tetrad is not needed to obtain any result in these notes; Eq. 2.52 already contains everything. What it buys is bookkeeping. Once the count is written as the rest-frame density n_g times a radial proper-length element $\mathrm{d}\ell$ times a transverse proper-area element $\boldsymbol{\sigma}$, each later contribution can be assigned to one of these three slots. The contraction $k \cdot u_g$ is already contained in $\mathrm{d}\ell$ and must not be counted again as a separate weight. Terms that seem to proliferate in a component calculation are often the same slot written in different variables.

[4] carries this construction to fully nonlinear order. The factorization survives, and the ruler distortions become the entries of an exact three by three matrix whose definition assumes only that the rulers are infinitesimal, not that the perturbations are small. Its determinant reproduces the third statement above: the mixed entries appear there only through $\mathcal{B}^2$ and $\mathcal{B}_m \mathcal{B}_{\bar{m}}$, never linearly. The treatment below is the first order of that expansion.

Convention when reading the source paper

[4] evaluates the spatial determinant at fixed source proper time and then multiplies by a separate clock Jacobian to convert to the observed redshift. That is the factored, comoving-ruler bookkeeping described in Eq. 3.33, not the fixed-physical-ruler convention these notes use. The two agree, but the individual $\mathcal{C}$ and $\mathcal{M}$ differ between them by terms proportional to $\mathcal{T}$. See the convention check in Sec. 3.6 before transcribing any formula.

2.10 Observed coordinates and the exact inferred density

The observer does not assign arbitrary coordinates to the light cone. A source is located by quantities measured at the detector: redshift and an angular direction, with the observation made at a specified observer proper time. These labels form a three-dimensional observed-coordinate manifold. A fiducial background cosmology is used only to assign an apparent spatial volume to a small cell in these labels.

Historical note: Ellis and observational cosmology

The insistence that cosmological data live on the observer's past light cone, and that a cosmological model should be formulated in terms of what is measurable there, was developed systematically by[8] and collaborators under the name observational cosmology. The programme asked which spacetime geometries are consistent with data on a single null cone, and it made explicit that the natural coordinates for the problem are the ones an observer actually reports. The construction in this section is a perturbative instance of that viewpoint: $(\tilde{z}, \tilde{\theta}, \tilde{\phi})$ are observer labels, and the fiducial cosmology enters only through the volume form assigned to a cell of them.

Fix the observer's proper time τ_o. On a regular patch of the past light cone Σ_{τ_o}, use the directly observed coordinates

$$y^a = (\tilde{z}, \tilde{\theta}, \tilde{\phi}). \tag{2.63}$$

The null geodesic and source worldline determine the physical embedding

$$\mathsf{X}_{\tau_o} : (\tilde{z}, \tilde{\theta}, \tilde{\phi}) \longmapsto X^\mu(\tilde{z}, \tilde{\theta}, \tilde{\phi}; \tau_o). \tag{2.64}$$

The fiducial FLRW relation defines

$$\tilde{a} = (1+\tilde{z})^{-1}, \qquad \tilde{\chi}(\tilde{z}) = \int_0^{\tilde{z}} \frac{\mathrm{d}z}{H(z)}. \tag{2.65}$$

The apparent physical volume form assigned to an observed cell is

$$\tilde{\boldsymbol{\omega}} \equiv \frac{\tilde{a}^3\tilde{\chi}^2}{H(\tilde{z})} \sin\tilde{\theta}\, \mathrm{d}\tilde{z} \wedge \mathrm{d}\tilde{\theta} \wedge \mathrm{d}\tilde{\phi}. \tag{2.66}$$

This is a form on the observed-coordinate manifold, not an induced volume form of the null hypersurface.

Changing the fiducial background changes the numerical volume assigned to a fixed observed cell and therefore changes the inferred overdensity field in the same way that an Alcock–Paczynski remapping changes a catalogue coordinate system. It does not change the integer number of sources in the cell. The physical count form and the fiducial volume form must therefore be kept conceptually distinct.

Define the inferred density n_g^{obs} by the equality of three-forms

$$\mathsf{X}^*_{\tau_o}\mathcal{J} \equiv n_g^{\mathrm{obs}}(\tilde{z}, \hat{\mathbf{n}})\, \tilde{\boldsymbol{\omega}}. \tag{2.67}$$

Using Eq. 2.52,

$$n_g^{\mathrm{obs}} = \frac{H(\tilde{z})}{\tilde{a}^3\tilde{\chi}^2 \sin\tilde{\theta}} \sqrt{-g}\, \epsilon_{\mu\nu\rho\sigma} n_g u_g^{\mu} \frac{\partial X^{\nu}}{\partial\tilde{z}} \frac{\partial X^{\rho}}{\partial\tilde{\theta}} \frac{\partial X^{\sigma}}{\partial\tilde{\phi}}. \tag{2.68}$$

The orientation is understood to make the right-hand side positive. Eq. 2.68 is nonperturbative and coordinate-independent, although its displayed components refer to a coordinate chart on spacetime.

Coordinate independence follows because both sides of Eq. 2.67 are three-forms on the same observed manifold. Their ratio is a scalar function of $(\tilde{z}, \hat{\mathbf{n}})$. The metric determinant, four-velocity, and embedding derivatives in Eq. 2.68 are individually coordinate-dependent components, but the full antisymmetric contraction is not.

The observed overdensity is then

$$1 + \delta_g^{\mathrm{obs}}(\tilde{z}, \hat{\mathbf{n}}) \equiv \frac{n_g^{\mathrm{obs}}(\tilde{z}, \hat{\mathbf{n}})}{\bar{n}_g(\tilde{z})}. \tag{2.69}$$

A selection function can be included by replacing j^{μ} with the current of the selected population or by multiplying the pulled-back count by a response to measured flux, size, color, and other observables.

This separation is useful because source selection has two aspects. A rest-frame cut changes the tracer population and hence its bias coefficients. A cut on an observed quantity can also respond to propagation effects such as magnification. The latter is an additional observable response and should not be absorbed into the rest-frame density bias.

2.11 Linearizing a pullback

Perturbing the observed count requires varying two ingredients at once. The physical current form differs from its FLRW value, and the event selected by a fixed observed redshift and angle is displaced from the background event. Keeping only one of these variations produces a gauge-dependent result. The Lie derivative provides a compact way to include the displacement of the embedding.

The exact expression becomes the usual perturbative observable after expanding both the current form and the light-cone embedding. Let

$$\mathcal{J} = \overline{\mathcal{J}} + \delta\mathcal{J}, \qquad X^\mu(y) = \overline{X}^\mu(y) + \Delta x^\mu(y). \tag{2.70}$$

To derive the linear expansion cleanly, extend Δx^μ off the background image of $\overline{\mathsf{X}}$ and let φ_λ be its flow. Locally,

$$\mathsf{X} = \varphi_1 \circ \overline{\mathsf{X}}. \tag{2.71}$$

For any differential form α,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\varphi_\lambda^*\alpha\bigg|_{\lambda=0} = \mathcal{L}_{\Delta x}\alpha. \tag{2.72}$$

Therefore

$$\begin{aligned} \mathsf{X}^*\mathcal{J} &= \overline{\mathsf{X}}^*\varphi_1^*\left(\overline{\mathcal{J}} + \delta\mathcal{J}\right) \\ &= \overline{\mathsf{X}}^*\left[\overline{\mathcal{J}} + \delta\mathcal{J} + \mathcal{L}_{\Delta x}\overline{\mathcal{J}}\right] + \mathcal{O}(2). \end{aligned} \tag{2.73}$$

The extension of Δx^μ away from the background light cone is only a device for defining a flow. At linear order the pulled-back result depends on the displacement along the embedded surface and on the background form evaluated there. Physically, the Lie-derivative term includes both the change in the background density along the displacement and the deformation of the coordinate cell. The two perturbative pieces have distinct meanings: $\delta\mathcal{J}$ changes the physical current at a fixed background event, while $\mathcal{L}_{\Delta x}\overline{\mathcal{J}}$ moves the source event selected by fixed observed coordinates.

Now apply the passive coordinate transformation Eq. 2.1, with which this chapter opened. The perturbation of a covariant form transforms as

$$\delta\mathcal{J} \mapsto \delta\mathcal{J} - \mathcal{L}_\xi\overline{\mathcal{J}}, \tag{2.74}$$

while the coordinate displacement of the same physical source event obeys $\Delta x \mapsto \Delta x + \xi$. Consequently,

$$\begin{aligned} &(\delta\mathcal{J} - \mathcal{L}_\xi\overline{\mathcal{J}}) + \mathcal{L}_{\Delta x+\xi}\overline{\mathcal{J}} \\ &= \delta\mathcal{J} + \mathcal{L}_{\Delta x}\overline{\mathcal{J}}. \end{aligned} \tag{2.75}$$

This is the geometric origin of the cancellation among source, redshift, volume, propagation, and observer terms in a gauge-invariant number-count formula.

A passive gauge transformation relabels the same source event. The perturbation of the current at a fixed coordinate point changes by $-\mathcal{L}_{\xi}\overline{\mathcal{J}}$, while the coordinate displacement to the physical event changes by $+\xi^{\mu}$. The opposite signs are not an accidental feature of a particular formula; they express the fact that changing the field at a fixed point and moving the point are complementary parts of the same pullback.

The gauge-invariant object is not a specially chosen density perturbation. It is the pulled-back current form divided by a volume form defined from observed coordinates. Gauge-dependent density and displacement perturbations are the two parts of the linear expansion of that one geometric object.

! Key idea: the pullback resolves the gauge ambiguity automatically

In the usual presentation one computes many separate contributions to the number count, finds that each is gauge dependent, and then observes that the total is not. The cancellation looks like a fortunate accident, and checking it is laborious. Eq. 2.75 shows that it is neither. The count was defined as the integral of one geometric object over one surface, so it could not have depended on coordinates; splitting it into a field perturbation and a displacement is a choice made by the calculator, and the two pieces must therefore transform oppositely. The practical consequence is a reliable bookkeeping rule. Any term that appears in $\delta\mathcal{J}$ has a partner in $\mathcal{L}_{\Delta x}\overline{\mathcal{J}}$, and a calculation that drops observer terms, or evaluates the source at the fiducial rather than the displaced event, has broken the pairing rather than found a new effect.

Looking back at the first half of the chapter with Eq. 2.75 in hand, the invariance of $\mathcal{T}$ in Eq. 2.38 is no longer a computation that happened to work. It is one instance of the general statement, restricted to the temporal factor of the count. The same will be true of the ruler distortions constructed in Chapter 3: each is a ratio of a measured quantity to an inferred one, and each is invariant for the same structural reason.

Summary

Key ideas. A density perturbation is meaningless until a slicing is named, because it compares the physical density at one event with a background value at another, and a gauge transformation changes which events are paired. The mismatch between the two slicings that matter, source proper time and observed redshift, is itself gauge invariant and is the cosmic clock $\mathcal{T}$. Behind every such invariance is one structural fact: a catalogue counts worldlines crossing a surface, the object that does this covariantly is the current three-form $\mathcal{J} = \iota_j\boldsymbol{\varepsilon}$, and a three-form can be pulled back to any three-manifold, including a past light cone that carries no induced volume element. The count is therefore the integral of one geometric object, and splitting it into a field perturbation $\delta\mathcal{J}$ and a displacement $\mathcal{L}_{\Delta x}\overline{\mathcal{J}}$ is a bookkeeping choice whose two halves must transform oppositely. Adapting the frame to the observation factorizes the count into a radial leg and a transverse area, which is why only two of the six ruler distortions can change a volume at linear order.

Main results.

- The passive gauge map Eq. 2.1 with scalar generator Eq. 2.2, and the resulting density transformations Eq. 2.5 and Eq. 2.6.
- The metric perturbation transformations Eq. 2.10 and Eq. 2.11, tabulated in Table C.1.
- The observer-to-source initial-value problem Eq. 2.20, which fixes the photon momentum

locally at the observer and uses the measured redshift to select the emission event.

- The displacement field Eq. 2.22, its transformation Eq. 2.23, and the redshift perturbation Eq. 2.27, including the observer term that enforces fixed observer proper time.
- The cosmic clock Eq. 2.37, its invariance Eq. 2.38, and the slicing relation Eq. 2.43 between δ_m^{or} and δ_m^{pt}.
- The current three-form Eq. 2.46, its conservation property Eq. 2.48, and the count as a pullback integral Eq. 2.50.
- The coordinate expression for the pullback, Eq. 2.52, valid on null and spacelike surfaces alike, and the spacelike check Eq. 2.54.
- The null-tetrad factorization Eq. 2.61 and count Eq. 2.62, separating the observable into rest-frame density, radial proper length, and transverse proper area without double-counting $k \cdot u_g$.
- The exact inferred density Eq. 2.68, defined by the ratio of the pulled-back count form to the fiducial volume form Eq. 2.66.
- The linearized pullback Eq. 2.73 and the cancellation Eq. 2.75.

What the next chapter builds on. The clock $\mathcal{T}$ supplies the temporal half of the map from a source density to an observed count. Chapter 3 constructs the spatial half by comparing a physical ruler at the source with the separation inferred from observed labels, decomposes the result into six distortions, and takes the determinant to obtain the volume Jacobian. Combining that determinant with $\mathcal{T}$ and the proper-time bias relation of Chapter 1 produces the central result of Part I.

3 Cosmic rulers, cosmic clocks, and the observed density

Where we are

Chapter 2 established two things. The observed count is the pullback of a current three-form to the observed-coordinate manifold, which makes it gauge invariant by construction. And the source proper time at which a bias response is defined differs from the background time assigned to the observed redshift, by the gauge-invariant amount $H^{-1}\mathcal{T}$. What remains is the spatial counterpart of that second statement: the volume of the cell into which the sources are binned is also not what the fiducial cosmology says it is.

This chapter supplies that piece and then assembles the result. The method is the same one used for the clock. Define a quantity operationally, in this case the ratio of a physical separation measured in the source rest frame to the separation an observer infers from redshifts and angles, and the ratio is automatically invariant even though every term in it is not.

A cosmic ruler is any small physical separation whose rest-frame length or statistical scale is known or modeled. The observer infers a separation from redshift and angle using a fiducial FLRW relation. Metric perturbations, velocities, and photon propagation change the relation between the physical and inferred separations. At linear order these changes form six independent ruler distortions, and their determinant gives the spatial volume part of the number-count Jacobian.

The six distortions are organized by how they sit relative to the line of sight: one longitudinal scalar $\mathcal{C}$, a two-component transverse vector $\mathcal{B}_i$, and a symmetric two-dimensional tensor $\mathcal{A}_{ij}$ whose trace $\mathcal{M}$ is the magnification and whose trace-free part is a shear. Only $\mathcal{C}$ and $\mathcal{M}$ can enter a scalar volume at linear order, and the chapter closes by combining them with the clock and the proper-time bias relation into

$$\delta_g^{\rm obs} = \delta_g^{\rm pt} + (b_e - 3)\mathcal{T} - \mathcal{C} - \mathcal{M},$$

which is Eq. 3.38 below and the central result of Part I. The Kaiser field of Chapter 1 is its subhorizon limit.

One warning is worth stating in advance, because it is the most common way to get this calculation wrong. The bookkeeping used here fixes physical rulers, so the ruler determinant already delivers the physical volume at the source event. An additional transported-volume factor $1 + 3\mathcal{T}$ does not belong on top of it. The alternative convention, in which such a factor is explicit, requires correspondingly shifted ruler distortions. Both are correct; mixing them double counts the clock.

Figure placeholder. The screen basis at an observed position. A short segment of the past light cone with the line of sight $\hat{\mathbf{n}}$, the two transverse screen directions, and three infinitesimal rulers drawn along them, annotated to show which distortion ($\mathcal{C}$, $\mathcal{B}_i$, $\mathcal{A}_{ij}$) deforms which pair of legs.

3.1 A physical ruler and its inferred separation

The construction is local at the source in the sense that the two endpoints are taken infinitesimally close compared with the scale over which the background and perturbations vary. It is nevertheless an observed light-cone quantity because the endpoint coordinates are inferred from photons reaching the observer. The screen projector separates radial and angular directions defined by the observed line of sight.

At an observed position $(\tilde{z}, \hat{\mathbf{n}})$, define

$$P_{ij} = \delta_{ij} - \hat{n}_i \hat{n}_j, \qquad V_\parallel = \hat{n}_i V^i, \qquad V_\perp^i = P^i{}_j V^j. \tag{3.1}$$

The derivatives used below are

$$\partial_\parallel = \hat{n}^i \partial_i, \qquad \partial_{\perp i} = P_i^{\ j} \partial_j, \qquad \partial_{\tilde{\chi}} = \partial_\parallel - \partial_\eta \tag{3.2}$$

along the background past light cone at fixed observed direction.

For two nearby observed source positions, the fiducial FLRW separation is

$$\tilde{r}^2 = \tilde{a}^2 \delta_{ij} \delta\tilde{x}^i \delta\tilde{x}^j, \qquad \tilde{r}_c^2 \equiv \delta_{ij} \delta\tilde{x}^i \delta\tilde{x}^j. \tag{3.3}$$

A physical ruler is defined by simultaneous separation in the local source rest frame:

$$\boxed{r_0^2 = (g_{\mu\nu} + u_\mu u_\nu) \delta x^\mu \delta x^\nu}. \tag{3.4}$$

This definition is covariant. The projector $g_{\mu\nu} + u_\mu u_\nu$ specifies the equal-proper-time comparison that is implicit in any physical length measurement.

Without the velocity projector, two endpoints that are simultaneous in the observer's background coordinates need not be simultaneous in the source rest frame. The $u_\mu u_\nu$ term removes the component of the separation along the source four-velocity and leaves the spatial length measured by a locally comoving observer.

The actual and inferred source positions satisfy

$$x^\mu(\tilde{x}) = \tilde{x}^\mu + \Delta x^\mu(\tilde{x}). \tag{3.5}$$

For endpoints separated by $\delta\tilde{x}^\mu$,

$$\begin{aligned} \delta x^\mu &= x^\mu(\tilde{x} + \delta\tilde{x}) - x^\mu(\tilde{x}) \\ &= \delta\tilde{x}^\mu + \delta\tilde{x}^\alpha \partial_\alpha \Delta x^\mu + \mathcal{O}(\delta\tilde{x}^2). \end{aligned} \tag{3.6}$$

The two inferred endpoints lie on the background light cone, so

$$\delta\tilde{x}^0 = -\delta\tilde{x}_\parallel. \tag{3.7}$$

It follows that

$$\begin{aligned}\delta\tilde{x}^\alpha\partial_\alpha &= -\delta\tilde{x}_\parallel\partial_\eta + (\delta\tilde{x}_\parallel\hat{n}^i + \delta\tilde{x}_\perp^i)\partial_i \\ &= \delta\tilde{x}_\parallel\partial_{\tilde{\chi}} + \delta\tilde{x}_\perp^i\partial_{\perp i}.\end{aligned} \tag{3.8}$$

Eq. 3.8 is the differential form of the light-cone constraint. A radial change in the inferred source position changes both the spatial radius and the emission time, whereas a transverse change moves the endpoint on the screen at fixed background distance. Derivatives of Δx^μ with respect to these directions are the local deformation tensor of the observed-to-physical map.

3.2 Expansion of the rest-frame length

The purpose of the following algebra is to express the covariant length Eq. 3.4 entirely in terms of perturbations evaluated at the fiducial observed position. There are three sources of the difference: the scale factor at emission is shifted, the local metric and velocity alter the rest-frame projector, and the displacement field changes the endpoint separation.

For the metric convention of Eq. 2.7, the source four-velocity is

$$u^\mu = a^{-1}(1 - A, v^i), \qquad u_\mu = a(-1 - A, v_i - B_i) \tag{3.9}$$

to linear order. Hence

$$\begin{aligned}(g+uu)_{00} &= 0,\\ (g+uu)_{0i} &= -a^2 v_i,\\ (g+uu)_{ij} &= a^2(\delta_{ij} + h_{ij}).\end{aligned} \tag{3.10}$$

The cancellation of A and B_i in these projector components is not a gauge choice. It follows from using the physical source four-velocity in the equal-proper-time projector. The lapse and shift still affect the observable through the redshift and displacement fields, but they do not appear as independent local length perturbations. The lapse and shift cancel from the local spatial projector except through the physical velocity. Substitution into Eq. 3.4 gives

$$r_0^2 = a^2\left[(\delta_{ij} + h_{ij})\delta x^i \delta x^j - 2v_i \delta x^0 \delta x^i\right]. \tag{3.11}$$

At the source event,

$$a(x^0) = \tilde{a}(1 + \Delta \ln a). \tag{3.12}$$

Use Eq. 3.6, Eq. 3.7, and Eq. 3.8 and keep one perturbation. The result is

$$\begin{aligned}\frac{r_0^2 - \tilde{r}^2}{\tilde{a}^2} &= 2\Delta\ln a\,\delta_{ij}\delta\tilde{x}^i\delta\tilde{x}^j + h_{ij}\delta\tilde{x}^i\delta\tilde{x}^j + 2v_i\delta\tilde{x}_\parallel\delta\tilde{x}^i \\ &\quad + 2\delta_{ij}\delta\tilde{x}^i\left(\delta\tilde{x}_\parallel\partial_{\tilde{\chi}} + \delta\tilde{x}_\perp^k\partial_{\perp k}\right)\Delta x^j.\end{aligned} \tag{3.13}$$

Because $r_0 = \tilde{r} + \mathcal{O}(1)$,

$$\frac{\tilde{r} - r_0}{\tilde{r}} = -\frac{r_0^2 - \tilde{r}^2}{2\tilde{r}^2}. \tag{3.14}$$

Eq. 3.13 is the algebraic source of all six linear ruler observables.

Each term has a direct origin. The first rescales all lengths because the actual emission scale factor differs from the fiducial one. The second is the local spatial metric distortion. The third enforces simultaneity in the moving source frame. The final term is the gradient of the light-cone displacement and contains radial redshift distortions, transverse deflection, and their mixed components.

3.3 Longitudinal, mixed, and transverse distortions

A symmetric deformation of three-dimensional separations has six components. Relative to the observed line of sight they decompose into one longitudinal scalar $\mathcal{C}$, a two-component transverse vector $\mathcal{B}_i$, and a symmetric two-dimensional tensor $\mathcal{A}_{ij}$ with three components. This decomposition is kinematic and does not assume scalar perturbations; vector and tensor metric perturbations fit into the same formulas.

Decompose the orientation dependence as[2]

$$\frac{\tilde{r}-r_0}{\tilde{r}} = \mathcal{C}\frac{(\delta\tilde{x}_\parallel)^2}{\tilde{r}_c^2} + \mathcal{B}_i\frac{\delta\tilde{x}_\parallel\delta\tilde{x}_\perp^i}{\tilde{r}_c^2} + \mathcal{A}_{ij}\frac{\delta\tilde{x}_\perp^i\delta\tilde{x}_\perp^j}{\tilde{r}_c^2}. \tag{3.15}$$

Collecting the coefficient of $(\delta\tilde{x}_\parallel)^2$ in Eq. 3.13 gives

$$\boxed{\mathcal{C} = -\Delta\ln a - \frac{1}{2}h_\parallel - v_\parallel - \partial_{\tilde{\chi}}\Delta x_\parallel}. \tag{3.16}$$

The mixed coefficient is

$$\boxed{\mathcal{B}_i = -P_i{}^j h_{jk}\hat{n}^k - v_{\perp i} - \hat{n}^k\partial_{\perp i}\Delta x_k - \partial_{\tilde{\chi}}\Delta x_{\perp i}}. \tag{3.17}$$

The transverse tensor is

$$\boxed{\mathcal{A}_{ij} = -\Delta\ln a\, P_{ij} - \frac{1}{2}P_i{}^k P_j{}^l h_{kl} - \frac{1}{2}\left(P_j{}^k\partial_{\perp i} + P_i{}^k\partial_{\perp j}\right)\Delta x_k}. \tag{3.18}$$

Since $\Delta x_i = \hat{n}_i\Delta x_\parallel + \Delta x_{\perp i}$ and the line-of-sight basis varies across the sky,

$$\mathcal{A}_{ij} = -\Delta\ln a\, P_{ij} - \frac{1}{2}P_i{}^k P_j{}^l h_{kl} - \partial_{\perp(i}\Delta x_{\perp j)} - \frac{\Delta x_\parallel}{\tilde{\chi}}P_{ij}. \tag{3.19}$$

The last term is the change in transverse physical scale caused by a radial displacement.

The three observables have different operational meanings. $\mathcal{C}$ changes a ruler aligned with the line of sight and contains the relativistic completion of redshift-space distortion. $\mathcal{B}_i$ mixes a radial and a transverse leg and describes a skewing of the inferred coordinate grid. $\mathcal{A}_{ij}$ acts within the screen and contains both isotropic magnification and spin-two shear.

The complete combinations $\mathcal{C}$, $\mathcal{B}_i$, and $\mathcal{A}_{ij}$ are gauge-invariant because both r_0 and $\tilde{r}$ are operationally defined. Their individual metric, velocity, and displacement pieces are not.

This is another instance of the pullback logic. A gauge transformation changes the coordinate displacement and the local metric evaluated at the fiducial endpoint, but it cannot change the ratio of the measured physical ruler to the ruler inferred from fixed observed labels.

! Key idea: clocks and rulers are operational, not auxiliary

$\mathcal{T}$, $\mathcal{C}$, $\mathcal{B}_i$, and $\mathcal{A}_{ij}$ are not convenient groupings of metric perturbations chosen to make an answer come out invariant. Each is defined as a comparison between two things an observer can in principle obtain: a physical duration or length in the source rest frame, and the duration or length that the fiducial cosmology assigns to the same pair of observed labels. Their invariance is a consequence of that definition, not a property to be checked afterwards. The practical payoff is that they can be reasoned about independently of the perturbation variables. One can ask what $\mathcal{M}$ does to a survey without first choosing a gauge, and one can identify a term as physical or as bookkeeping by asking which measurement it changes. The same logic, applied to a local rather than a light-cone measurement, produces the Fermi frame of Chapter 4.

3.4 Magnification and ruler shear

The transverse tensor is naturally separated into a trace and a trace-free part. The trace changes the apparent area of a small source and therefore affects angular diameter distance and number-count dilution. The trace-free part changes shape at fixed area and is the relativistic ruler shear.

The trace and trace-free parts of the transverse distortion are

$$
\begin{aligned}
\mathcal{M} &\equiv P^{ij}\mathcal{A}_{ij}, \\
\gamma_{ij}^{\rm ruler} &\equiv \mathcal{A}_{ij} - \frac{1}{2}P_{ij}\mathcal{M}.
\end{aligned}
\tag{3.20}
$$

Define the coordinate convergence

$$
\hat{\kappa} \equiv -\frac{1}{2}\partial_{\perp i}\Delta x^i_\perp . \tag{3.21}
$$

Taking the trace of Eq. 3.19 gives

$$
\boxed{\mathcal{M} = -2\Delta\ln a - \frac{1}{2}(h^i{}_i - h_\parallel) + 2\hat{\kappa} - 2\frac{\Delta x_\parallel}{\tilde{\chi}}}. \tag{3.22}
$$

The trace-free part is

$$
\begin{aligned}
\gamma_{ij}^{\rm ruler} = &-\frac{1}{2}\left(P_i{}^k P_j{}^l - \frac{1}{2}P_{ij}P^{kl}\right)h_{kl} \\
&- \partial_{\perp(i}\Delta x_{\perp j)} - P_{ij}\hat{\kappa}.
\end{aligned}
\tag{3.23}
$$

The coordinate convergence $\hat{\kappa}$ is not an observable by itself. Local metric terms and radial/scale-factor perturbations complete the observable.

In the standard weak-lensing limit the additional terms are suppressed or cancel against endpoint conventions, leaving the familiar line-of-sight convergence. On horizon scales or in a general gauge, however, identifying $\hat{\kappa}$ alone with magnification omits pieces required by the operational distance measurement.

The angular- and luminosity-distance perturbations are

$$\frac{\Delta D_A}{D_A} = \frac{\Delta D_L}{D_L} = -\frac{1}{2}\mathcal{M}, \tag{3.24}$$

where the equality follows from distance duality. In the subhorizon weak-lensing limit, $\mathcal{M} \to 2\kappa$.

3.5 The ruler determinant

Number counts require the volume change rather than the distortion of one chosen ruler. Three independent infinitesimal separation vectors span a small rest-frame volume. The determinant of their linear map gives the ratio between that physical volume and the volume inferred from observed coordinates.

Choose an orthonormal apparent basis $(\hat{e}_\parallel, \hat{e}_1, \hat{e}_2)$ and write the unit orientation of a ruler as $\hat{q}^a$. The fractional length distortion has the form

$$\frac{\tilde{r} - r_0}{\tilde{r}} = \mathsf{D}_{ab}\hat{q}^a\hat{q}^b, \tag{3.25}$$

with symmetric part

$$\mathsf{D} = \begin{pmatrix} \mathcal{C} & \mathcal{B}_1/2 & \mathcal{B}_2/2 \\ \mathcal{B}_1/2 & \mathcal{A}_{11} & \mathcal{A}_{12} \\ \mathcal{B}_2/2 & \mathcal{A}_{12} & \mathcal{A}_{22} \end{pmatrix}. \tag{3.26}$$

An unobservable infinitesimal rotation may be added to the vector map, but it neither changes ruler lengths nor contributes to the determinant at linear order. Thus one may write

$$r_0^a = (\delta^a{}_b - \mathsf{D}^a{}_b)\tilde{r}^b \tag{3.27}$$

for the volume calculation. Three independent infinitesimal rulers span volumes related by

$$\begin{aligned} \frac{V_0}{\widetilde{V}} &= \det(I - \mathsf{D}) \\ &= 1 - \operatorname{tr}\mathsf{D} + \mathcal{O}(2) \\ &= 1 - \mathcal{C} - \mathcal{M} + \mathcal{O}(2). \end{aligned} \tag{3.28}$$

The mixed ruler $\mathcal{B}_i$ cannot enter a scalar volume at linear order. It first appears quadratically through the determinant.

This follows from rotational structure as well as from the determinant. A single transverse vector cannot form a scalar without another transverse vector. At linear order the volume responds only to the radial dilation and the transverse trace, $\mathcal{C} + \mathcal{M}$.

3.6 Adding the clock and the source density

The current three-form combines the physical source density with the physical volume element through which the source worldlines pass. For the ruler convention used above, the determinant in Eq. 3.28 already maps an observed coordinate cell to the rest-frame physical volume at the actual source event. This point is important: the ruler was defined with the source projector $g_{\mu\nu} + u_\mu u_\nu$, so the volume obtained from $\mathcal{C}$ and $\mathcal{M}$ is not a background volume at the fiducial time. It is the local volume measured on the source rest-frame hypersurface at the emission event.

Let $\delta_g^{\rm or}$ denote the rest-frame physical density perturbation evaluated at the source event but compared with the background mean at the time assigned to the observed redshift:

$$n_g(x_s) = \bar{n}_g(\tilde{a}) \left[1 + \delta_g^{\rm or}(\tilde{z}, \hat{\mathbf{n}})\right] . \tag{3.29}$$

Since $V_0/\widetilde{V} = 1 - \mathcal{C} - \mathcal{M}$ is already the physical rest-frame volume ratio at the source event, the number of sources in the observed cell is

$$1 + \delta_g^{\rm obs} = (1 + \delta_g^{\rm or})(1 - \mathcal{C} - \mathcal{M}) + \mathcal{O}(2). \tag{3.30}$$

Thus, for the fixed-physical-ruler convention of Eq. 3.4,

$$\boxed{\delta_g^{\rm obs} = \delta_g^{\rm or} - \mathcal{C} - \mathcal{M}}. \tag{3.31}$$

There is a true background identity

$$\frac{V_0(t_F)}{V_0(\bar{t}(\tilde{a}))} = 1 + 3\mathcal{T} \tag{3.32}$$

for a physical volume transported with the background source congruence between two nearby proper times. It should not be multiplied into Eq. 3.30 when $\mathcal{C}$ and $\mathcal{M}$ are the fixed-physical-ruler distortions derived above. Doing so would count the same background volume expansion twice. Equivalently, one may introduce evolving or comoving rulers and keep an explicit $1 + 3\mathcal{T}$ factor; then the ruler distortions themselves must be shifted by the corresponding time-evolution terms. The invariant number count is the same, but the bookkeeping differs.

In the notation of Schmidt–Jeong cosmic rulers, an evolving ruler has extra terms proportional to $\mathrm{d}\ln r_0/\mathrm{d}\ln a$ in the longitudinal and transverse distortions. With the sign conventions of this chapter, for a comoving ruler these shifts are

$$\mathcal{C}_{\rm com} = \mathcal{C} + \mathcal{T}, \qquad \mathcal{A}_{ij}^{\rm com} = \mathcal{A}_{ij} + \mathcal{T} P_{ij}, \qquad \mathcal{M}_{\rm com} = \mathcal{M} + 2\mathcal{T}, \tag{3.33}$$

so that, to linear order,

$$(1 + 3\mathcal{T})(1 - \mathcal{C}_{\rm com} - \mathcal{M}_{\rm com}) = 1 - \mathcal{C} - \mathcal{M}. \tag{3.34}$$

This is the convention used in factored current-form presentations such as[4]. The present notes keep the fixed-physical-ruler convention because it follows directly from Eq. 3.4 and from the determinant in Sec. 3.5.

The bias response is naturally defined at fixed source proper time. Write

$$n_g(x_s) = \bar{n}_g(a_F)[1 + \delta_g^{\rm pt}(x_s)]. \tag{3.35}$$

Since $\ln(a_F/\tilde{a}) = \mathcal{T}$,

$$\bar{n}_g(a_F) = \bar{n}_g(\tilde{a}) \left[1 + \frac{\mathrm{d}\ln \bar{n}_g}{\mathrm{d}\ln a}\mathcal{T}\right], \tag{3.36}$$

and hence

$$\delta_g^{\rm or} = \delta_g^{\rm pt} + \frac{\mathrm{d}\ln \bar{n}_g}{\mathrm{d}\ln a}\mathcal{T} = \delta_g^{\rm pt} + (b_e - 3)\mathcal{T}. \tag{3.37}$$

The combination $b_e - 3$ is the logarithmic derivative of the physical mean density, while b_e is the derivative of the comoving mean abundance. Substitution into Eq. 3.31 gives

$$\boxed{\delta_g^{\rm obs}(\tilde{z}, \hat{\mathbf{n}}) = \delta_g^{\rm pt} + (b_e - 3)\mathcal{T} - \mathcal{C} - \mathcal{M}} \tag{3.38}$$

for a volume-limited sample at linear order. At leading bias order,

$$\delta_g^{\rm pt} = b_1 \delta_m^{\rm pt}, \tag{3.39}$$

so

$$\boxed{\delta_g^{\rm obs} = b_1 \delta_m^{\rm pt} + (b_e - 3)\mathcal{T} - \mathcal{C} - \mathcal{M}}. \tag{3.40}$$

The gauge issue is resolved by specifying all objects operationally: the source density at fixed proper time, the clock relating that time to the observed redshift, and the radial and transverse ruler distortions of the observed volume element.

For a conserved comoving population, $b_e = 0$, so the physical mean density evolves as a^{-3} and the clock contribution is $-3\mathcal{T}$. This term is not cancelled by an additional transported-volume factor in the present convention, because the physical volume at the source event has already been supplied by the ruler determinant. For a population whose comoving abundance evolves, b_e adds the genuine selection-history response.

Convention check

Eq. 3.38 uses fixed physical rulers. If instead one factors the current-form Jacobian into a clock factor and a spatial determinant at constant proper time, the spatial determinant must be built from the corresponding evolving-ruler distortions, as in Eq. 3.33. Mixing the fixed-physical-ruler $\mathcal{C}, \mathcal{M}$ with an extra $1 + 3\mathcal{T}$ factor double counts the clock.

Sample selection

Eq. 3.38 is the geometric count for a volume-limited population. Flux, color, size, or other cuts define additional responses of the selected current. They must be included separately and are not part of the rest-frame bias relation.

3.7 Flux selection and the lensing coefficient

For a threshold sample, define

$$s \equiv \frac{\mathrm{d}\log_{10}\bar{N}(< m)}{\mathrm{d}m}. \tag{3.41}$$

In the weak-lensing limit, the transverse area increase gives the dilution -2κ through $-\mathcal{M}$. Magnification raises the flux of sources across the threshold and contributes $+5s\kappa$. Therefore

$$\Delta_g^{\mathrm{lens}} = (5s-2)\kappa. \tag{3.42}$$

The coefficient is the sum of a geometric area Jacobian and a sample-selection response.

The two contributions should be distinguished when generalizing the sample definition. The -2κ term is present for any count per observed solid angle because the same sources are spread over a changed apparent area. The $5s\kappa$ term depends on how rapidly the selected abundance changes at the flux threshold and vanishes for a genuinely volume-limited sample.

3.8 Newtonian limit

For $k \gg \mathcal{H}$, in the plane-parallel weak-field limit,

$$\mathcal{C} \longrightarrow \frac{1}{\mathcal{H}}\partial_\parallel v_\parallel, \qquad \mathcal{M} \longrightarrow 2\kappa. \tag{3.43}$$

The clock and potential terms in ordinary equal-redshift clustering are suppressed by powers of $\mathcal{H}/k$ relative to density and the velocity gradient. Neglecting lensing and selection,

$$\delta_g^{\mathrm{obs}} \longrightarrow b_1\delta_m - \frac{1}{\mathcal{H}}\partial_\parallel v_\parallel, \tag{3.44}$$

which is the Kaiser field of Chapter 1. The Newtonian formula is the short-distance limit of the same covariant observable, not a different definition of the catalogue field.

The hierarchy of terms can be understood dimensionally. The density and velocity gradient scale as order unity in $\mathcal{H}/k$, a velocity or potential-gradient redshift term is typically suppressed by $\mathcal{H}/k$, and a potential term by $(\mathcal{H}/k)^2$, modulo line-of-sight integrals and selection coefficients. These suppressions justify the Newtonian approximation on short scales but do not alter the observable definition.

3.9 Complete linear galaxy-clustering prescription

The preceding sections can now be assembled into a single prescription. For pressureless matter, synchronous-comoving gauge is the constant-matter-proper-time slicing, so

$$\delta_m^{\mathrm{pt}} = \delta_m^{\mathrm{sc}}. \tag{3.45}$$

For a volume-limited sample, the complete linear observable is therefore

$$\boxed{\Delta_g^{\rm obs}(\tilde z,\hat{\mathbf n}) = b_1\delta_m^{\rm sc} + (b_e-3)\mathcal{T} - \mathcal{C} - \mathcal{M}}. \tag{3.46}$$

Here the four terms have distinct origins: $b_1\delta_m^{\rm sc}$ is the local source response at fixed proper time, $\mathcal{T}$ moves that source response to the event selected at fixed observed redshift, and $\mathcal{C}$ and $\mathcal{M}$ are the radial and transverse Jacobians of the observed-to-physical coordinate map. A general sample adds a selection response $\Delta_{\rm sel}$:

$$\Delta_g^{\rm obs} = b_1\delta_m^{\rm sc} + (b_e-3)\mathcal{T} - \mathcal{C} - \mathcal{M} + \Delta_{\rm sel}. \tag{3.47}$$

For a flux threshold, $\Delta_{\rm sel}$ contains the magnification response $5s\kappa$; the geometric area dilution is already contained in $-\mathcal{M}$. Other color, size, or detection cuts require their own response coefficients.

Substituting the definitions of the clock and ruler distortions gives a useful component form valid in any gauge:

$$\boxed{\begin{aligned}\Delta_g^{\rm obs} = {}& b_1\delta_m^{\rm sc} + b_e\Delta\ln a + (b_e-3)HI_A + \frac{1}{2}h^i{}_i + v_\parallel + \partial_{\tilde\chi}\Delta x_\parallel \\ & - 2\hat\kappa + 2\frac{\Delta x_\parallel}{\tilde\chi} + \Delta_{\rm sel}.\end{aligned}} \tag{3.48}$$

Although the separate terms on the right-hand side depend on the gauge used to perform the calculation, the displayed combination does not. The density is written as $\delta_m^{\rm sc}$ because that is the gauge-invariant proper-time density entering the bias relation; all remaining fields may be evaluated in whichever gauge is convenient.

In synchronous-comoving gauge, $A = B = v = 0$ and $I_A = 0$, so $\mathcal{T} = \Delta\ln a$. The same result becomes

$$\boxed{\begin{aligned}\Delta_g^{\rm obs} = {}& b_1\delta_m^{\rm sc} + b_e(\Delta\ln a)_{\rm sc} + \frac{1}{2}(h^i{}_i)_{\rm sc} + \partial_{\tilde\chi}(\Delta x_\parallel)_{\rm sc} \\ & - 2\hat\kappa_{\rm sc} + 2\frac{(\Delta x_\parallel)_{\rm sc}}{\tilde\chi} + \Delta_{\rm sel}.\end{aligned}} \tag{3.49}$$

The redshift perturbation in this gauge follows directly from the photon-energy equation,

$$(\Delta\ln a)_{\rm sc} = \frac{1}{2}\int_0^{\tilde\chi} \mathrm{d}\chi\,(h'_\parallel)_{\rm sc}, \tag{3.50}$$

up to the chosen observer monopole convention. The radial displacement and convergence are obtained by integrating the spatial null-geodesic equations with the observer boundary data of Sec. 2.3. Thus Eq. 3.49 is a closed algorithm: solve the linear Einstein–matter system in synchronous-comoving gauge, ray trace from the observer, and insert the resulting endpoint and line-of-sight fields.

For a single growing adiabatic mode, every linear quantity in Eq. 3.46 is a linear functional of $\delta_m^{\rm sc}$. It is useful to define transfer kernels by

$$\mathcal{T} = \mathcal{R}_{\mathcal{T}}\,\delta_m^{\rm sc}, \qquad \mathcal{C} = \mathcal{R}_{\mathcal{C}}\,\delta_m^{\rm sc}, \qquad \mathcal{M} = \mathcal{R}_{\mathcal{M}}\,\delta_m^{\rm sc}, \qquad \Delta_{\rm sel} = \mathcal{R}_{\rm sel}\,\delta_m^{\rm sc}, \tag{3.51}$$

where the kernels include endpoint terms and line-of-sight integrals and therefore depend on k, direction, and redshift. The master equation is then

$$\boxed{\Delta_g^{\rm obs}(\mathbf{k},\hat{\mathbf{n}},\tilde{z}) = \left[b_1 + (b_e - 3)\mathcal{R}_{\mathcal{T}} - \mathcal{R}_{\mathcal{C}} - \mathcal{R}_{\mathcal{M}} + \mathcal{R}_{\rm sel}\right]\delta_m^{\rm sc}(\mathbf{k},\tilde{z})}. \tag{3.52}$$

Eq. 3.52 should be understood as a transfer-function statement, not as a local multiplication in configuration space. In particular, lensing and integrated redshift terms sample the matter perturbation along the entire line of sight. Once the cosmological model fixes the linear transfer functions, however, $\delta_m^{\rm sc}$ supplies the single stochastic initial field and the bracket supplies the complete deterministic map to the observed galaxy fluctuation.

! The completed observable chain

The calculation has now closed the chain

$$\delta_m^{\rm sc} \longrightarrow \text{local tracer response} \longrightarrow \text{clock and ruler map} \longrightarrow \Delta_g^{\rm obs}(\tilde{z},\hat{\mathbf{n}}).$$

Gauge freedom changes the intermediate representation, but not this map.

Summary

Key ideas. A ruler distortion is the ratio of a physical rest-frame separation, defined with the projector $g_{\mu\nu} + u_\mu u_\nu$ so that the comparison is made at equal proper time, to the separation inferred from observed labels. Because both ends of that ratio are operational, the six distortions are gauge invariant while their metric, velocity, and displacement pieces are not. Only the radial dilation $\mathcal{C}$ and the transverse trace $\mathcal{M}$ can enter a scalar volume at linear order, since a single transverse vector cannot form a scalar. Finally, the observed count is the product of a source density and a volume ratio, and the convention chosen for the ruler determines which of the two carries the background expansion.

Main results.

- The covariant physical ruler Eq. 3.4 and the master expansion Eq. 3.13 from which all six distortions follow.
- The longitudinal, mixed, and transverse distortions Eq. 3.16, Eq. 3.17, and Eq. 3.18, with the transparent form Eq. 3.19.
- Magnification Eq. 3.22, ruler shear Eq. 3.23, and the distance perturbation Eq. 3.24.
- The volume Jacobian Eq. 3.28, in which $\mathcal{B}_i$ is absent at linear order.
- The central result Eq. 3.38, its biased form Eq. 3.40, and the complete prescriptions Eq. 3.46 and Eq. 3.52.
- The flux-selection coefficient Eq. 3.42, separating the geometric dilution -2κ from the selection response $5s\kappa$.
- The Newtonian limit Eq. 3.44, recovering Chapter 1.

What the next chapter builds on. Part I is now complete: the observed count has been expressed in terms of a proper-time source density, a clock, and two ruler distortions, all of them operationally defined on the past light cone. Chapter 4 turns to the complementary question. Instead of comparing a source with an observer across a finite null geodesic, it asks what a single freely falling observer can measure inside a small laboratory. The answer, curvature rather than connection, supplies the missing justification for the operator basis assumed in Chapter 1 and sets up the tidal physics of Chapter 5.

Part II

Part II: Local Physics

4 Fermi normal coordinates: the Manasse–Misner construction

Where we are

Part I produced observables of a single kind. Every one of them mixes source physics, propagation, and observer terms that acquire meaning only in combination, which is why establishing gauge invariance took the machinery of the last two chapters. Part II turns to measurements that need none of it.

The preceding observables compare a source with an observer across a finite null geodesic and are therefore intrinsically nonlocal. We now ask a different question: what gravitational information can be measured inside a small freely falling laboratory? Fermi normal coordinates answer this by constructing coordinates around an entire timelike geodesic. The construction removes the connection on the central worldline while retaining the curvature that produces relative acceleration.

Two debts from Part I are settled here. Chapter 1 asserted that the local gravitational operators available to a formation process are the density and the trace-free tide, on the grounds that a freely falling observer can discard a constant potential and a uniform acceleration. That was a Newtonian argument. This chapter replaces it with a theorem: there exist coordinates in which the connection vanishes along an entire geodesic, the metric is Minkowski to first order in the spatial distance from it, and the leading correction is quadratic with coefficients fixed by the Riemann tensor. Whatever a local experiment can measure must therefore be built from curvature. Separately, the chapter supplies the frame in which Chapter 5 defines an intrinsic galaxy shape, which must be a tensor in a local orthonormal rest frame rather than in a cosmological chart.

The chapter opens by fixing two conventions, the curvature sign and the meaning of a hatted index, and then, in Sec. 4.2, by asking what such coordinates could possibly achieve. Counting the freedom in a coordinate change against the number of metric coefficients at each order settles both how far the gravitational field can be removed and what is guaranteed to survive. The construction then proceeds in four steps: fix a tetrad on a central geodesic and parallel transport it; use the exponential map along orthogonal spacelike geodesics to assign coordinates to nearby events; show that the resulting chart is regular near the worldline; and Taylor expand the metric, using the geodesic identities to convert connection derivatives into curvature. No field equation is used at any point, so the result holds in any smooth Lorentzian spacetime.

Figure placeholder. Construction of Fermi normal coordinates. The central timelike geodesic $\mathcal{G}$ with the parallel-transported tetrad drawn at two values of proper time; from one of them, a fan of spacelike geodesics orthogonal to $e_{\hat{0}}$ sweeping out the constant-$X^{\hat{0}}$ surface; a nearby event P labelled by the proper time τ of its originating point and the proper distance s along the geodesic reaching it.

4.1 Conventions

Two conventions are fixed here and then used without further comment: the sign and index order of the Riemann tensor, and what a hat on an index means.

4.1.1 Curvature

We use

$$\begin{aligned} R^{\rho}{}_{\sigma\mu\nu} &= \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma}, \\ [\nabla_\mu, \nabla_\nu] V^\rho &= R^{\rho}{}_{\sigma\mu\nu} V^\sigma. \end{aligned} \tag{4.1}$$

Fixing the curvature convention is essential because the sign of the quadratic Fermi metric and of geodesic deviation changes when the Riemann tensor is defined with the opposite commutator order. The formulas below are internally tied to Eq. 4.1; translating a result from another reference requires translating both its index order and its curvature sign.

> **Caveat**
>
> Manasse and Misner write the first covariant index before the contravariant index, $R_\mu{}^\nu{}_{\alpha\beta}$. Their mixed-index differential definition agrees with Eq. 4.1 after translating the index order. However, their all-lowered object $R^{\mathrm{MM}}_{\mu\nu\alpha\beta}$ has the first pair reversed relative to ours, so $R^{\mathrm{MM}}_{\mu\nu\alpha\beta} = -R_{\mu\nu\alpha\beta}$. This accounts for the opposite signs between their printed Eq. (66) and Eq. 4.48, Eq. 4.49, and Eq. 4.50 below.

4.1.2 Frame indices

Two kinds of index appear from here on, and the distinction matters enough to state before it is used.

An unhatted Greek index μ, ν, ρ, σ, as in Eq. 4.1, labels a component in an arbitrary chart on spacetime. Nothing is assumed about that chart. A hatted index labels a component in the Fermi coordinate chart $X^{\hat{a}}$.

On the central geodesic the Fermi coordinate basis is constructed to coincide with the parallel-transported orthonormal tetrad $e_{\hat{a}}$, which is Eq. 4.24. Hence hatted components of a **tensor evaluated on** $\mathcal{G}$ also equal its tetrad projections. Away from the worldline they are Fermi-coordinate components and should not be read as local Lorentz components. This distinction matters for the metric expansion itself, whose hatted components are functions of X^i away from $\mathcal{G}$.

Within this chapter the notation is:

Index	Range	Meaning
μ, ν, ρ, σ, and μ' for a second chart	0 to 3	arbitrary chart on spacetime; never a frame index
$\hat{a}, \hat{b}, \hat{c}, \hat{d}$	0 to 3	Fermi-coordinate components; on $\mathcal{G}$, also observer-tetrad components

Index	Range	Meaning
$\hat{0}$		Fermi time direction; on $\mathcal{G}$, the observer-tetrad time direction
i, j, k, l, m	1 to 3	spatial Fermi-coordinate directions within this chapter; hats are suppressed

Within this chapter, Latin letters denote spatial Fermi-coordinate indices, so the hat is suppressed. Every four-dimensional Fermi index carries a hat, including the time direction, which is $\hat{0}$. In Chapter 5, named weak-field and TT charts also use Latin letters for their Cartesian spatial coordinates; there a bare time index 0 identifies a chart component, while $\hat{0}$ identifies the Fermi-frame quantity measured by the observer.

Two consequences are worth spelling out. The coordinates are $X^{\hat{a}} = (X^{\hat{0}}, X^i)$ and the tetrad is $e_{\hat{a}} = (e_{\hat{0}}, e_i)$, so the exponential map Eq. 4.17 pairs X^i with e_i and the notation matches the structure. And $\Gamma^{\hat{a}}_{ij}$, to take the case that most often provokes the question, is a Christoffel symbol of the Fermi chart whose two lower indices have been restricted to spatial directions while the upper one still runs over all four; the hat on $\hat{a}$ and its absence on i, j say exactly that.

Where a bare 0 does appear

Chapter 5 writes R_{0i0j} with a bare time index in two places, and this is deliberate rather than a lapse. Both occur inside a named chart rather than in the observer frame: the weak-field metric of Eq. 5.1, written in local physical coordinates, and the transverse-traceless chart of Eq. 5.15. The frame quantities in that chapter, $\delta R_{\hat{0}i\hat{0}j}$ and $\mathcal{E}^{\mathrm{GW}}_{ij} = R^{\mathrm{GW}}_{\hat{0}i\hat{0}j}$, carry the hat as they should. Reading the hat is therefore enough to tell a chart component from a measured one.

4.2 The goal, and a degree-of-freedom count

Before constructing anything it is worth stating precisely what is wanted and asking whether it can be had.

We want coordinates adapted to one freely falling observer. Erect an orthonormal tetrad at a point of the observer's worldline, thread coordinates outward from it, and then expand the metric in the resulting chart. The aim is to remove as much of the gravitational field as the coordinates permit, and to find out what is left when no more can be removed. The leftover is the physics, because it is what no observer can transform away.

The question is how far the removal can go. Write the map from an arbitrary chart x^μ to the coordinates $X^{\hat{a}}$ we are building as a Taylor series about the origin,

$$x^\mu = A^\mu{}_{\hat{a}} X^{\hat{a}} + \frac{1}{2!} B^\mu{}_{\hat{a}\hat{b}} X^{\hat{a}} X^{\hat{b}} + \frac{1}{3!} C^\mu{}_{\hat{a}\hat{b}\hat{c}} X^{\hat{a}} X^{\hat{b}} X^{\hat{c}} + \mathcal{O}(X^4), \tag{4.2}$$

with B symmetric in its two lower indices and C totally symmetric in its three. The metric in the new chart is likewise a series,

$$g_{\hat a\hat b}(X) = g_{\hat a\hat b}(0) + g_{\hat a\hat b,\hat c}(0)X^{\hat c} + \frac{1}{2}g_{\hat a\hat b,\hat c\hat d}(0)X^{\hat c}X^{\hat d} + \mathcal{O}(X^3). \tag{4.3}$$

At each order the coefficients of Eq. 4.2 are the freedom, and the coefficients of Eq. 4.3 are what we are trying to set to zero. Counting both in $n = 4$ spacetime dimensions settles the matter.

Order	Metric coefficients	Available freedom	Balance
$g_{\hat a\hat b}$	$n(n+1)/2 = 10$	$A^\mu{}_{\hat a}$: $n^2 = 16$	6 left over
$g_{\hat a\hat b,\hat c}$	$10n = 40$	$B^\mu{}_{\hat a\hat b}$: 40	exactly enough
$g_{\hat a\hat b,\hat c\hat d}$	$10 \times 10 = 100$	$C^\mu{}_{\hat a\hat b\hat c}$: 80	20 cannot be removed

A counting argument is only as good as the structure that licenses it, so it is worth saying why the rows can be read independently. The system is triangular. The value $g_{\hat a\hat b}(0)$ involves A alone; the first derivatives involve A and B, with B entering linearly; the second derivatives involve A, B, and C, with C entering linearly. One may therefore solve order by order, each order fixing the next coefficient while the earlier ones are already determined and merely contribute inhomogeneous terms. What matters at each order is not a vague supply of adjustable parameters but whether one specific linear map is onto.

Read the rows in turn.

At zeroth order there is more freedom than there are conditions. Ten conditions fix the metric to $\eta_{\hat a\hat b}$, and the six surviving parameters are the Lorentz group: having chosen one orthonormal frame, a boost or a rotation gives another.

At first order the count is tight, and the reason is worth making explicit rather than leaving to the arithmetic. Define the Jacobian and its inverse by

$$A^\mu{}_{\hat a} \equiv \left.\frac{\partial x^\mu}{\partial X^{\hat a}}\right|_0, \qquad A^{\hat a}{}_\mu A^\mu{}_{\hat b} = \delta^{\hat a}{}_{\hat b}, \qquad A^\mu{}_{\hat a}A^{\hat a}{}_\nu = \delta^\mu{}_\nu. \tag{4.4}$$

The hatted Christoffel symbol is a coordinate connection coefficient of the Fermi chart,

$$\nabla_{\partial_{\hat b}}\partial_{\hat c} = \Gamma^{\hat a}{}_{\hat b\hat c}\partial_{\hat a}, \qquad \Gamma^{\hat a}{}_{\hat b\hat c} = \frac{1}{2}g^{\hat a\hat d}(\partial_{\hat b}g_{\hat c\hat d} + \partial_{\hat c}g_{\hat b\hat d} - \partial_{\hat d}g_{\hat b\hat c}). \tag{4.5}$$

It is not the tetrad projection of the old Christoffel symbols, because a connection is not a tensor. At the origin its transformation law is

$$\Gamma^{\hat a}{}_{\hat b\hat c}(0) = A^{\hat a}{}_\rho\left[B^\rho{}_{\hat b\hat c} + \Gamma^\rho{}_{\mu\nu}(0)A^\mu{}_{\hat b}A^\nu{}_{\hat c}\right]. \tag{4.6}$$

The Hessian B is the inhomogeneous part. Therefore demanding $\Gamma^{\hat a}{}_{\hat b\hat c}(0) = 0$ determines it uniquely:

$$\boxed{B^\rho{}_{\hat a\hat b} = -A^\mu{}_{\hat a}A^\nu{}_{\hat b}\,\Gamma^\rho_{\mu\nu}(0)}. \tag{4.7}$$

Here $\Gamma^\rho_{\mu\nu}$ carries Greek indices throughout: it is the ordinary spacetime connection of the original chart x^μ, evaluated at the origin, and not a tetrad-frame quantity. The two factors of

A convert its lower spacetime indices into the hatted indices carried by B; the inverse Jacobian in Eq. 4.6 converts the upper index of the new connection.

The same result follows directly from the metric. Differentiating $g_{\hat{a}\hat{b}} = g_{\mu\nu}(\partial_{\hat{a}}x^{\mu})(\partial_{\hat{b}}x^{\nu})$ and taking the cyclic combination $g_{\hat{a}\hat{b},\hat{c}} + g_{\hat{a}\hat{c},\hat{b}} - g_{\hat{b}\hat{c},\hat{a}}$ gives

$$0 = 2g_{\mu\nu}A^{\mu}{}_{\hat{a}}\left[B^{\nu}{}_{\hat{b}\hat{c}} + \Gamma^{\nu}{}_{\rho\sigma}A^{\rho}{}_{\hat{b}}A^{\sigma}{}_{\hat{c}}\right], \tag{4.8}$$

so nondegeneracy of $g_{\mu\nu}A^{\mu}{}_{\hat{a}}$ reproduces Eq. 4.7. The local-inertial map through quadratic order is consequently

$$x^{\mu} = z^{\mu} + A^{\mu}{}_{\hat{a}}X^{\hat{a}} - \frac{1}{2}\Gamma^{\mu}{}_{\nu\rho}(z)A^{\nu}{}_{\hat{a}}A^{\rho}{}_{\hat{b}}X^{\hat{a}}X^{\hat{b}} + \mathcal{O}(X^3). \tag{4.9}$$

B is therefore not an independent supply of forty knobs that happens to match forty conditions; it is solved for, uniquely, in terms of A and the spacetime connection at the origin. The equality of the two counts is the shadow of a sharper fact: a connection and the inhomogeneous term $B^{\rho}{}_{\hat{a}\hat{b}}$ carry the same index symmetry, symmetric in the lower pair, so the map from B to the connection it induces is a bijection. That is why the first derivatives can always be removed, and why nothing is left over once they are. This is the equivalence principle as a statement about linear algebra: a freely falling observer detects no gravitational field through the connection because the connection is pure inhomogeneous term.

At second order the map is no longer onto. The coefficient $C^{\mu}{}_{\hat{a}\hat{b}\hat{c}}$ is totally symmetric in its three lower indices, while $g_{\hat{a}\hat{b},\hat{c}\hat{d}}$ is symmetric within the pair $\hat{a}\hat{b}$ and within the pair $\hat{c}\hat{d}$ but has no symmetry relating the two pairs. The two objects sit in different symmetry classes, and the shift that C can generate spans only an eighty-dimensional subspace of the hundred-dimensional target. The twenty directions outside that image are unreachable by any choice of coordinates, and they are precisely the combinations carrying the algebraic symmetries of the Riemann tensor. Twenty is not a coincidence: it is $n^2(n^2-1)/12$ evaluated at $n = 4$.

Two ways of trying to evade this fail, and seeing why confirms that the twenty are genuine. The six Lorentz parameters left unspent at zeroth order act on the surviving components as a linear representation; a group action moves a nonzero tensor around its orbit but never to the origin, so the residual frame freedom rotates the twenty rather than reducing them. And one might try spending B differently, declining to remove all forty first derivatives in the hope of buying something at second order. That trades away forty quantities a local experiment would detect for at most twenty it cannot remove anyway, which is never worth doing.

! Key idea: the counting tells us where to stop and what we will find

The three rows answer both questions at once. They say that $g_{\hat{a}\hat{b},\hat{c}} = 0$ is the right target, because it is the last condition that can be met completely. And they say that the expansion must therefore be carried to X^2, because that is the first order at which anything survives, and that what survives there is exactly the Riemann tensor. The quadratic term of the Fermi metric is not one contribution among several; at that order it is the whole of the local gravitational field.

There remains the question the counting above does not answer. It was carried out at a single point, which gives Riemann normal coordinates. We want the conditions to hold not at one event but at every proper time along the observer's worldline, which is infinitely many copies of the same demand. Is that still possible?

It is, and the reason is that two independent geometric prescriptions divide the work exactly between them. Along the central geodesic the connection has the same forty components, which split according to how many of the two lower indices are spatial:

Components	Count	Removed by
$\Gamma^{\hat{a}}_{ij}$	$4 \times 6 = 24$	the coordinate lines being spatial geodesics
$\Gamma^{\hat{b}}_{\hat{0}\hat{a}}$	$4 \times 4 = 16$	parallel transport of the tetrad

The first block is killed by threading the coordinates outward along geodesics that leave the worldline orthogonally, so that each radial coordinate line has vanishing acceleration. The second is killed by carrying the tetrad along the worldline without rotating it, which also covers $\Gamma^{\hat{a}}_{\hat{0}\hat{0}} = 0$, the statement that the central curve is itself a geodesic. Together they account for all forty, with none left over and none demanded twice.

So the construction is available, and the remainder of the chapter carries it out. Sec. 4.4 builds the map by the exponential prescription, Sec. 4.6 verifies the two blocks in the table, and Sec. 4.9 extracts the twenty surviving components at order X^2.

Two caveats belong with the count. It is local: nothing here guarantees that the coordinates remain single valued far from the worldline, and they do not. And it assumes the central worldline is a geodesic. Allowing acceleration removes the geodesic condition that supplied $\Gamma^{\hat{a}}_{\hat{0}\hat{0}} = 0$, and a term linear in X^i reappears in $g_{\hat{0}\hat{0}}$.

4.3 Geometric data on the central geodesic

The central geodesic represents the ideal freely falling observer. A coordinate system around this observer also requires a prescription for spatial axes. Parallel transport supplies nonaccelerating, nonrotating axes along a geodesic and allows tensor components measured at different proper times to be compared in one tetrad frame.

> **i** Historical note: from Fermi to Manasse and Misner
>
> The idea of erecting coordinates along a worldline so that the metric is Minkowski and its first derivatives vanish all along it, rather than at a single event as in Riemann normal coordinates, goes back to[9]. Fermi's construction addressed the general question of what happens in the neighbourhood of a timelike curve; the extension to accelerated and rotating observers, now standard in relativistic metrology, uses Fermi–Walker rather than parallel transport. The systematic quadratic expansion used in this chapter, with the metric coefficients written in terms of the Riemann tensor on the central geodesic, is due to[5], who also introduced the affine-rescaling argument reproduced in Eq. 4.18. Their index conventions differ from the ones used here, which is the source of an apparent sign discrepancy; Appendix D carries out the translation.

Choose a point P_0 and a future-directed unit timelike vector $e_{\hat{0}}(0)$ at P_0. Let

$$z : \tau \longmapsto z(\tau) \tag{4.10}$$

be the timelike geodesic with

$$z(0) = P_0, \qquad \dot{z}^\mu = e^\mu_{\hat{0}}, \qquad \nabla_{e_{\hat{0}}} e_{\hat{0}} = 0, \qquad e_{\hat{0}} \cdot e_{\hat{0}} = -1. \tag{4.11}$$

Choose three spacelike vectors $e_i(0)$ at P_0 so that

$$e_{\hat{a}}(0) \cdot e_{\hat{b}}(0) = \eta_{\hat{a}\hat{b}}. \tag{4.12}$$

Parallel transport the full tetrad along $z(\tau)$:

$$\nabla_{e_{\hat{0}}} e_{\hat{a}} = 0. \tag{4.13}$$

Metric compatibility preserves orthonormality for all τ.

Because the tetrad is orthonormal, $e_{\hat{0}}$ is the observer's four-velocity and the three e_i span its instantaneous rest space. Because it is parallel transported, no inertial acceleration or rotation is introduced by the frame. If the central worldline were accelerated, Fermi–Walker rather than parallel transport would be required and linear acceleration terms would remain in the local metric.

The construction depends on the choice of the initial spatial triad only through a constant spatial rotation. Choosing a different central geodesic defines a different local observer.

4.4 Orthogonal geodesics and the coordinate map

At each proper time on the central worldline, shoot spacelike geodesics in all directions orthogonal to $e_{\hat{0}}$. The initial tangent specifies a spatial displacement in the local tetrad. The exponential map sends that tangent vector to the endpoint of the corresponding geodesic. This construction defines both simultaneity and spatial position relative to the observer.

Following Secs. II and IV of[5], let

$$h(\tau; \alpha^i; \lambda) \tag{4.14}$$

denote the spacelike geodesic with affine parameter λ satisfying

$$\begin{aligned} &h(\tau; \alpha; 0) = z(\tau), \\ &\left.\frac{\partial h}{\partial \lambda}\right|_{\lambda=0} = \alpha^i e_i(\tau), \qquad \delta_{ij}\alpha^i \alpha^j = 1. \end{aligned} \tag{4.15}$$

The initial tangent is orthogonal to the central four-velocity. For a point with proposed Fermi coordinates $(X^{\hat{0}}, X^i)$, define

$$s \equiv (\delta_{ij} X^i X^j)^{1/2}, \qquad \alpha^i \equiv \frac{X^i}{s} \tag{4.16}$$

for $s \neq 0$. The physical point is

$$P(X) = h(X^{\hat{0}}; \alpha^i; s) = \exp_{z(X^{\hat{0}})}\left(X^i e_i(X^{\hat{0}})\right). \tag{4.17}$$

Thus $X^{\hat{0}} = \tau$ is proper time on the central geodesic, while s is proper distance along the orthogonal spacelike geodesic at its initial point.

The statement about s is local to the initial point: affine parametrization preserves the norm of the tangent along the spacelike geodesic, so its parameter length equals proper length. Surfaces of constant $X^{\hat{0}}$ are generated by these orthogonal geodesics; away from the central worldline they need not be orthogonal to a global timelike congruence.

Manasse and Misner use the affine-rescaling identity

$$h(\tau; s\alpha^i; \lambda) = h(\tau; \alpha^i; s\lambda). \tag{4.18}$$

Both sides solve the same geodesic equation with the same initial point and tangent after the parameter rescaling, so uniqueness proves the identity. Eq. 4.17 can therefore be written without separating length and direction:

$$P(X) = h(X^{\hat{0}}; X^i; 1). \tag{4.19}$$

This is the form used to establish differentiability and regularity at $X^i = 0$.

Writing the map directly in terms of X^i avoids the apparent singularity of the direction $\alpha^i = X^i/s$ at the origin. Smooth dependence of geodesic solutions on their initial data makes $h(\tau; X^i; 1)$ a smooth function of the Cartesian tangent-space components.

4.5 Regularity near the central geodesic

A geometric prescription does not automatically guarantee a valid coordinate chart. The Jacobian must be nonsingular, and each nearby event must be reached by a unique orthogonal geodesic from the central worldline. The first property follows locally from the tetrad; the second restricts the construction to a normal tubular neighborhood.

Let $y^{\mu'}$ be any regular coordinates in a neighborhood of the central geodesic. The coordinate transformation is

$$y^{\mu'}(X) = h^{\mu'}(X^{\hat{0}}; X^i; 1). \tag{4.20}$$

From the initial tangent of the orthogonal geodesics,

$$\left.\frac{\partial y^{\mu'}}{\partial X^i}\right|_{X^j=0} = e_i^{\mu'}(X^{\hat{0}}). \tag{4.21}$$

Along $X^i = 0$, $P(X) = z(X^{\hat{0}})$, so

$$\left.\frac{\partial y^{\mu'}}{\partial X^{\hat{0}}}\right|_{X^j=0} = e_{\hat{0}}^{\mu'}(X^{\hat{0}}). \tag{4.22}$$

Therefore

$$\det\left.\left(\frac{\partial y^{\mu'}}{\partial X^{\hat{a}}}\right)\right|_{\mathcal{G}} = \det(e_{\hat{a}}^{\mu'}) \neq 0, \tag{4.23}$$

where $\mathcal{G}$ denotes the central geodesic. The inverse-function theorem guarantees a nonsingular coordinate system in a sufficiently small tubular neighborhood. Globally the construction fails when orthogonal geodesics intersect or reach conjugate points; FNC are intrinsically local.

The failure at a caustic is geometric rather than perturbative: two different spatial tangent vectors would be assigned to the same event, so the Fermi coordinates cease to be one-to-one. For cosmological applications the laboratory size is chosen well below both the curvature radius and the scale at which such intersections occur.

4.6 The Fermi conditions

Riemann normal coordinates make the connection vanish at one event. Fermi normal coordinates extend the same local-inertial property along the full central geodesic. The extension is possible because the temporal basis is the geodesic tangent and the spatial basis is parallel transported, while the off-worldline coordinate curves are themselves geodesics at fixed central proper time.

On $\mathcal{G}$, the Fermi coordinate basis is the transported tetrad:

$$\left.\frac{\partial}{\partial X^{\hat{a}}}\right|_{\mathcal{G}} = e_{\hat{a}}. \tag{4.24}$$

Hence

$$\boxed{g_{\hat{a}\hat{b}}|_{\mathcal{G}} = \eta_{\hat{a}\hat{b}}}. \tag{4.25}$$

For fixed $X^{\hat{0}} = \tau$ and fixed direction α^i, the coordinate curve

$$X^{\hat{0}} = \tau, \qquad X^i = \lambda\alpha^i \tag{4.26}$$

is precisely one of the orthogonal geodesics. Its coordinate acceleration vanishes, so the geodesic equation gives

$$\Gamma^{\hat{a}}_{ij}(\tau, \lambda\alpha)\alpha^i\alpha^j = 0. \tag{4.27}$$

At $\lambda = 0$, the direction α^i is arbitrary and the connection is symmetric in i, j. Therefore

$$\Gamma^{\hat{a}}_{ij}|_{\mathcal{G}} = 0. \tag{4.28}$$

The coordinate basis vectors on $\mathcal{G}$ are parallel transported by construction. Thus

$$0 = \nabla_{e_{\hat{0}}} e_{\hat{a}} = \Gamma^{\hat{b}}_{\hat{0}\hat{a}}|_{\mathcal{G}} \frac{\partial}{\partial X^{\hat{b}}}, \tag{4.29}$$

which implies

$$\Gamma^{\hat{b}}_{\hat{0}\hat{a}}|_{\mathcal{G}} = 0. \tag{4.30}$$

Together with torsion freedom and Eq. 4.28,

$$\boxed{\Gamma^{\hat{c}}_{\hat{a}\hat{b}}|_{\mathcal{G}} = 0}, \qquad \boxed{\partial_{\hat{c}} g_{\hat{a}\hat{b}}|_{\mathcal{G}} = 0}. \tag{4.31}$$

The second statement follows from metric compatibility. These are the Fermi conditions: the metric is rectangular and has no linear spatial dependence along the chosen geodesic.

The vanishing of first derivatives means that a pointlike experiment on the central worldline cannot detect a gravitational field through the connection. It does not imply that the metric is constant in a finite neighborhood. Second spatial derivatives remain and are fixed by curvature, so an extended experiment can measure differential acceleration.

! Key idea: the equivalence principle removes the connection, not the curvature

Eq. 4.31 is the precise content of the equivalence principle for an extended experiment. What can be transformed away is $\Gamma^{\hat{c}}_{\hat{a}\hat{b}}$, and it can be removed not merely at one event, as in Riemann normal coordinates, but along the entire worldline. What cannot be removed is the second derivative of the metric, because it contains the Riemann tensor, and a tensor that vanishes in one chart vanishes in all of them. This is why the gravitational field available to a local experiment begins at tidal order, and why the size of the apparatus enters: the dimensionless effect scales as RL^2, so a pointlike detector measures nothing at all. Every local operator used later in these notes, from the tidal bias of Chapter 1 to the intrinsic alignment and gravitational-wave response of Chapter 5, follows from this single statement.

4.7 Geodesic deviation in the convention of these notes

Geodesic deviation gives the operational meaning of the curvature components that enter the Fermi metric. The deviation vector ξ^μ joins neighboring geodesics at equal values of their affine parameter. Its second covariant derivative measures relative rather than absolute acceleration, so it is invariant under the removal of a uniform gravitational field.

Consider a two-parameter family of geodesics $P(\lambda, \sigma)$. Define

$$u \equiv \frac{\partial}{\partial \lambda}, \qquad \xi \equiv \frac{\partial}{\partial \sigma}. \tag{4.32}$$

Because (λ, σ) are coordinates on the two-surface swept out by the family,

$$[u, \xi] = 0, \qquad \nabla_u \xi = \nabla_\xi u. \tag{4.33}$$

The second equality follows from the first because the connection is torsion free, so the antisymmetric part of ∇u acting on ξ reduces to the Lie bracket.

Each λ curve is geodesic, $\nabla_u u = 0$. Therefore

$$\begin{aligned} \nabla_u \nabla_u \xi &= \nabla_u \nabla_\xi u \\ &= \nabla_\xi \nabla_u u + R(u, \xi) u \\ &= R(u, \xi) u. \end{aligned} \tag{4.34}$$

In components, using Eq. 4.1,

$$\boxed{\frac{D^2\xi^\mu}{D\lambda^2} = -R^\mu{}_{\nu\alpha\beta} u^\nu \xi^\alpha u^\beta}. \tag{4.35}$$

The minus-sign form follows by using antisymmetry in the last two curvature indices. For neighboring freely falling particles near the central timelike geodesic,

$$\frac{\mathrm{d}^2 X^i}{\mathrm{d}T^2} = -R^i{}_{\hat{0}j\hat{0}}(T) X^j + \mathcal{O}(X^2, \dot{X}X). \tag{4.36}$$

The matrix $R^i{}_{\hat{0}j\hat{0}}$ is the tidal acceleration per unit separation measured by the central observer. Its trace changes the volume of a small cloud of test particles, while its trace-free part shears the cloud. This same tensor will be used to describe scalar large-scale tides and gravitational waves.

4.8 Connection derivatives on the central geodesic

Since the connection itself vanishes on the central geodesic, the leading spatial variation of the connection is its first derivative. Curvature is precisely the antisymmetrized derivative of the connection at that location. The additional identities supplied by the radial geodesics determine the symmetric combinations needed for the Taylor expansion of the metric.

The quadratic metric requires first spatial derivatives of the connection. The Fermi conditions determine them algebraically from curvature, as in Sec. VII of[5].

Since $\Gamma^{\hat{a}}_{\hat{b}\hat{0}} = 0$ on $\mathcal{G}$ for every $X^{\hat{0}}$,

$$\Gamma^{\hat{a}}_{\hat{b}\hat{0},\hat{0}}|_{\mathcal{G}} = 0. \tag{4.37}$$

At $\mathcal{G}$, where quadratic connection terms vanish, the curvature definition gives

$$R^{\hat{a}}{}_{\hat{b}i\hat{0}} = \Gamma^{\hat{a}}_{\hat{b}\hat{0},i} - \Gamma^{\hat{a}}_{\hat{b}i,\hat{0}}. \tag{4.38}$$

The second term vanishes because $\Gamma^{\hat{a}}_{\hat{b}i} = 0$ along the full central geodesic. Hence

$$\boxed{\Gamma^{\hat{a}}_{\hat{b}\hat{0},i}|_{\mathcal{G}} = R^{\hat{a}}{}_{\hat{b}i\hat{0}}|_{\mathcal{G}}}. \tag{4.39}$$

Differentiate the radial identity Eq. 4.27 with respect to λ at the origin:

$$\Gamma^{\hat{a}}_{(ij,k)}|_{\mathcal{G}} = 0. \tag{4.40}$$

At $\mathcal{G}$,

$$R^{\hat{a}}{}_{ijk} = \Gamma^{\hat{a}}_{ik,j} - \Gamma^{\hat{a}}_{ij,k}. \tag{4.41}$$

Writing the analogous equation with i and j exchanged and using Eq. 4.40 gives

$$\boxed{\Gamma^{\hat{a}}_{ij,k}|_{\mathcal{G}} = -\frac{1}{3}\left(R^{\hat{a}}{}_{ijk} + R^{\hat{a}}{}_{jik}\right)_{\mathcal{G}}}. \tag{4.42}$$

The cyclic identity of the Riemann tensor is equivalent to the triple symmetry required by the radial geodesics.

Eq. 4.39 and Eq. 4.42 are the technical core of the Manasse–Misner construction. They convert a coordinate statement about the geodesic spray into covariant curvature coefficients. No field equation is used; the result holds in any smooth Lorentzian spacetime.

4.9 Quadratic expansion of the Fermi metric

The Fermi conditions remove the constant correction and all terms linear in the spatial coordinates. The first nontrivial metric coefficients are therefore quadratic in X^i. Metric compatibility relates their coefficients to the connection derivatives just obtained, and hence to the Riemann tensor on the central geodesic.

Metric compatibility gives

$$g_{\hat{a}\hat{b},\hat{c}} = g_{\hat{a}\hat{d}}\Gamma^{\hat{d}}_{\hat{b}\hat{c}} + g_{\hat{b}\hat{d}}\Gamma^{\hat{d}}_{\hat{a}\hat{c}}. \tag{4.43}$$

Differentiating with respect to X^i and evaluating on $\mathcal{G}$ yields

$$g_{\hat{a}\hat{b},ij}|_{\mathcal{G}} = \eta_{\hat{a}\hat{d}}\Gamma^{\hat{d}}_{\hat{b}i,j}|_{\mathcal{G}} + \eta_{\hat{b}\hat{d}}\Gamma^{\hat{d}}_{\hat{a}i,j}|_{\mathcal{G}}. \tag{4.44}$$

For the time–time component, Eq. 4.39 gives

$$\begin{aligned} g_{\hat{0}\hat{0},ij}|_{\mathcal{G}} &= -2\Gamma^{\hat{0}}_{\hat{0}i,j}|_{\mathcal{G}} \\ &= -2R_{\hat{0}i\hat{0}j}|_{\mathcal{G}}. \end{aligned} \tag{4.45}$$

For the mixed and spatial components, using Eq. 4.39 and Eq. 4.42 and the Riemann symmetries gives

$$\begin{aligned} g_{\hat{0}k,ij}|_{\mathcal{G}} &= -\frac{2}{3}\left(R_{\hat{0}ikj} + R_{\hat{0}jki}\right)_{\mathcal{G}}, \\ g_{\ell m,ij}|_{\mathcal{G}} &= -\frac{1}{3}\left(R_{\ell imj} + R_{\ell jmi}\right)_{\mathcal{G}}. \end{aligned} \tag{4.46}$$

The Taylor series

$$g_{\hat{a}\hat{b}}(T,\mathbf{X}) = \eta_{\hat{a}\hat{b}} + \frac{1}{2}g_{\hat{a}\hat{b},ij}(T,\mathbf{0})X^iX^j + \mathcal{O}(X^3) \tag{4.47}$$

then gives the Manasse–Misner metric in our all-lowered curvature convention:

$$\boxed{g_{\hat{0}\hat{0}} = -1 - R_{\hat{0}i\hat{0}j}(T)X^iX^j + \mathcal{O}(X^3)}. \tag{4.48}$$

$$\boxed{g_{\hat{0}i} = -\frac{2}{3}R_{\hat{0}jik}(T)X^jX^k + \mathcal{O}(X^3)}. \tag{4.49}$$

$$\boxed{g_{ij} = \delta_{ij} - \frac{1}{3}R_{ikjl}(T)X^kX^l + \mathcal{O}(X^3)}. \tag{4.50}$$

Curvature is evaluated on the central geodesic at the same proper time $T = X^{\hat{0}}$. Cubic terms contain covariant derivatives of curvature; their coefficients retain arbitrary proper-time dependence along the central geodesic.

The three equations encode different parts of local gravity. $g_{\hat{0}\hat{0}}$ controls the leading tidal acceleration of slow particles. $g_{\hat{0}i}$ contains gravitomagnetic-type curvature and affects moving particles and clock transport. g_{ij} describes the spatial curvature of the local coordinate grid. All are fixed by the same spacetime Riemann tensor rather than by independent potentials.

The mixed and spatial coefficients require a short manipulation using pair exchange and the first Bianchi identity, which is carried out in Appendix D. That appendix also translates the index order of[5], whose printed metric carries the opposite sign for the reason given in the caveat above.

As a check, the slow-particle geodesic equation from Eq. 4.48 gives

$$\Gamma^{i}_{\hat{0}\hat{0}} = R^{i}{}_{\hat{0}j\hat{0}} X^{j} + \mathcal{O}(X^{2}), \qquad \ddot{X}^{i} = -R^{i}{}_{\hat{0}j\hat{0}} X^{j}, \tag{4.51}$$

which agrees with geodesic deviation.

4.10 What is locally measurable?

The equivalence principle removes gravitational effects that can be represented by one value of the connection along a freely falling worldline. A finite-size experiment compares neighboring worldlines and is sensitive to derivatives of that connection, namely curvature. The operator content of a local effective description must therefore be expressible in terms of curvature, locally measured matter fields, and covariant derivatives or histories of these quantities.

The connection can be set to zero on $\mathcal{G}$, but curvature cannot. The electric part of the Riemann tensor seen by the observer is

$$\mathcal{R}_{ij} \equiv R_{\hat{0}i\hat{0}j}. \tag{4.52}$$

Its trace and trace-free parts are

$$\mathcal{R}_{ij} = \frac{1}{3}\delta_{ij}\mathcal{R}^{k}{}_{k} + \mathcal{R}_{\langle ij\rangle}, \qquad \mathcal{R}_{\langle ij\rangle} \equiv \left(\delta_{i}{}^{k}\delta_{j}{}^{l} - \frac{1}{3}\delta_{ij}\delta^{kl}\right)\mathcal{R}_{kl}. \tag{4.53}$$

The trace contributes isotropic geodesic focusing; the trace-free part is the shear tide. In vacuum it equals the electric Weyl tensor. In matter, the Ricci part carries local density and pressure information.

This is the relativistic completion of the Newtonian operator statement in Chapter 1. A local formation process cannot depend on a coordinate potential or a uniform acceleration. It can depend on curvature, its covariant derivatives, matter fields measured in the local tetrad, and their history along the worldline.

For nonrelativistic scalar perturbations the trace is related to the local density through the field equations, and the trace-free part reduces to the Newtonian tidal tensor. The Fermi construction explains why these are the leading gravitational operators in the bias expansion without relying on a particular cosmological gauge.

Key idea: Fermi normal coordinates isolate the genuine local observables

The value of the construction is not that it produces a convenient chart. It is that it settles which quantities a local process is permitted to depend on. A galaxy forming in a long-wavelength environment cannot know the value of Φ, because that value can be changed by a coordinate transformation that no local measurement can detect; it cannot know $\partial_i \Phi$ either, since a uniform acceleration is indistinguishable from free fall. It can respond to $\mathcal{R}_{ij}$, to matter fields measured in its own tetrad, and to the histories of both along its worldline. That is the entire list, and it is fixed by geometry rather than by modelling taste. The bias expansion of Chapter 1 was written down on the strength of a Newtonian version of this argument; here it is derived, and the derivation is what allows the same operator basis to be used in a relativistic setting where no preferred gauge is available.

4.11 Domain of validity

For a laboratory of characteristic size L, the quadratic approximation requires

$$|R_{\hat{a}\hat{b}\hat{c}\hat{d}}|L^2 \ll 1, \qquad |\nabla_{\hat{e}} R_{\hat{a}\hat{b}\hat{c}\hat{d}}|L^3 \ll 1. \tag{4.54}$$

The coordinate construction additionally requires a normal tubular neighborhood without intersections of orthogonal geodesics. FNC describe a local experiment around one timelike observer; they do not replace the past-light-cone coordinates required for cosmological observations over horizon-scale distances.

The conditions in Eq. 4.54 also state the derivative expansion explicitly. The dimensionless effect of curvature across the apparatus is RL^2; variation of the tide across the apparatus begins at ∇RL^3. A long-wavelength perturbation can therefore be treated as a nearly homogeneous local tide even when its metric amplitude is not itself a local observable.

Summary

Key ideas. Fermi normal coordinates extend the local-inertial property of Riemann normal coordinates from a single event to an entire timelike geodesic. Their construction uses only parallel transport and the exponential map, so it is available in any smooth Lorentzian spacetime and invokes no field equation. Along the central worldline the metric is $\eta_{\hat{a}\hat{b}}$ and the connection vanishes, which is the equivalence principle stated exactly; the first surviving structure is quadratic in the spatial coordinates, with coefficients built from the Riemann tensor. The chart is intrinsically local, failing where orthogonal geodesics intersect.

Counting settles in advance both where the construction must stop and what it will find. The forty first derivatives of the metric are matched by exactly forty coefficients in a coordinate change, so they can always be removed and nothing is left over; the one hundred second derivatives face only eighty, so twenty survive, and twenty is the number of independent Riemann components in four dimensions. Along a worldline rather than at a point the same forty connection components split into twenty-four killed by the radial geodesics and sixteen killed by parallel transport, again with nothing left over.

Main results.

- The degree-of-freedom count of Sec. 4.2, based on the chart expansion Eq. 4.2 and the metric expansion Eq. 4.3.
- The exponential-map definition Eq. 4.17 and its regular form Eq. 4.19, with regularity established by Eq. 4.23.
- The Fermi conditions Eq. 4.31: vanishing connection and vanishing first metric derivatives along $\mathcal{G}$.
- Geodesic deviation Eq. 4.35, and its nonrelativistic form Eq. 4.36 in which $R^{i}{}_{\hat{0}j\hat{0}}$ is the tidal acceleration per unit separation.
- The connection derivatives Eq. 4.39 and Eq. 4.42, which are the technical core of the construction.
- The quadratic Fermi metric Eq. 4.48, Eq. 4.49, and Eq. 4.50, and the consistency check Eq. 4.51.
- The electric Riemann tensor Eq. 4.52 and its trace split Eq. 4.53, giving isotropic focusing and shear tide.
- The domain of validity Eq. 4.54.

What the next chapter builds on. The chapter has identified the object a local experiment can measure: $\mathcal{R}_{ij} = R_{\hat{0}i\hat{0}j}$, evaluated in the observer's tetrad. Chapter 5 shows that three effects usually presented separately are the same object in different regimes. Its subhorizon scalar limit is the tidal operator K_{ij} assumed in Chapter 1; its trace-free part drives the intrinsic shape response that produces intrinsic alignment; and a gravitational wave contributes to it through $-\frac{1}{2}\ddot{h}_{ij}^{\mathrm{TT}}$. The chapter closes by contrasting this local response with the propagation distortion of Chapter 3, which can produce the same spin-two signal on the sky by an entirely different mechanism.

5 Local tides, intrinsic alignment, and gravitational waves

Where we are

Chapter 4 established that the gravitational information available inside a small freely falling laboratory is the electric part of the Riemann tensor, $\mathcal{R}_{ij} = R_{\hat{0}i\hat{0}j}$, measured in the observer's own tetrad. This chapter collects the consequences.

The local tensor $R_{\hat{0}i\hat{0}j}$ provides a common language for several effects that are often presented separately. Its scalar-perturbation limit is the tidal operator used in galaxy bias. A galaxy shape can respond to its trace-free part during formation, producing intrinsic alignment. A gravitational wave supplies a tensor contribution to the same local curvature and produces the familiar differential motion of freely falling particles.

The organizing observation is that a local source responds to a tensor representation, not to a mechanism. The response kernel cannot ask whether a trace-free tide was generated by scalar density perturbations or by a passing tensor mode; it sees only $\mathcal{E}_{\langle ij\rangle}$ and its history along the worldline. What distinguishes the two cases is the time dependence of the tide and its correlation with other observables, not the response itself.

Three results follow in sequence. First, the scalar limit of $\mathcal{R}_{ij}$ reproduces the tidal operator K_{ij} that Chapter 1 introduced on Newtonian grounds, which closes the argument begun there. Second, extending the functional-response formalism of Chapter 1 to a tensor-valued observable produces intrinsic alignment, with rotational invariance fixing the index structure and leaving a single scalar kernel R_E that carries the population-dependent internal dynamics. Third, the same construction applied to a gravitational wave shows why the local forcing is $\ddot{h}_{ij}^{\mathrm{TT}}$ rather than h_{ij}^{TT}.

The chapter ends with a distinction that matters for any survey measuring shapes. An observed spin-two field on the sky receives both a local source response, defined here, and the propagation distortion of Chapter 3. The two are indistinguishable by tensor structure alone and must be separated by where the response is defined.

5.1 The scalar tidal field in a local frame

We first recover the Newtonian tide from curvature. The metric in this subsection is written in local physical coordinates rather than in the conformal-coordinate convention of Chapter 2. The calculation isolates the perturbative Riemann tensor before restoring the FLRW scale-factor conversion.

Consider a weak-field metric in local physical coordinates,

$$\mathrm{d}s^2 = -(1+2\Psi)\mathrm{d}t^2 + (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{5.1}$$

where the sign convention for Φ is now the common local weak-field convention. The linearized Riemann tensor about Minkowski space is

$$R^{(1)}_{\rho\sigma\mu\nu} = \frac{1}{2}(\partial_\mu\partial_\sigma h_{\rho\nu} + \partial_\nu\partial_\rho h_{\sigma\mu} - \partial_\nu\partial_\sigma h_{\rho\mu} - \partial_\mu\partial_\rho h_{\sigma\nu}). \tag{5.2}$$

For Eq. 5.1,

$$R_{0i0j} = \partial_i\partial_j\Psi + \delta_{ij}\ddot{\Phi}. \tag{5.3}$$

In the quasistatic Newtonian limit, the time-derivative term is negligible and

$$R_{0i0j} \simeq \partial_i\partial_j\Psi. \tag{5.4}$$

The geodesic-deviation equation then becomes

$$\ddot{X}^i = -\partial^i\partial_j\Psi\, X^j, \tag{5.5}$$

which is the relative acceleration obtained by subtracting the force at the central worldline from the force at a neighboring point.

A constant Ψ produces no acceleration, and a term linear in position produces the same acceleration for every particle in the laboratory. Both disappear from relative motion. The Hessian is the first part of the potential that survives this subtraction, in agreement with the Fermi-coordinate expansion.

In cosmology, subtract the isotropic FLRW background and evaluate the perturbation in the observer's tetrad. On subhorizon scales,

$$\delta R_{\hat{0}i\hat{0}j} \simeq \frac{1}{a^2}\partial_i\partial_j\Psi. \tag{5.6}$$

The trace-free scalar tide is

$$\begin{aligned}\mathcal{E}^S_{ij} &\equiv \delta R_{\hat{0}\langle i|\hat{0}|j\rangle} \\ &\simeq \frac{1}{a^2}\left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\Psi.\end{aligned} \tag{5.7}$$

Using the Poisson equation,

$$\boxed{\mathcal{E}^S_{ij} \simeq \frac{3}{2}H^2\Omega_m(a)K_{ij}} \tag{5.8}$$

with K_{ij} defined in Eq. 1.18. Thus the Newtonian tidal operator in the bias expansion is the subhorizon limit of a local curvature component.

Here $\Omega_m = \Omega_m(a)$. The proportionality factor depends on time through $H^2\Omega_m$, while K_{ij} is dimensionless by definition. One may formulate a response using either K_{ij} or the dimensionful curvature $\mathcal{E}^S_{ij}$, but the corresponding bias coefficient has a different normalization. The functional definition must specify which operator is held fixed.

5.2 Intrinsic shape as a tensor-valued response

An intrinsic shape is a property of the source before its photons propagate to the observer. Unlike the scalar abundance, it carries two spatial indices in the source rest frame and can therefore respond linearly to a trace-free tide. The response formalism used for density bias extends directly once the free tensor indices and the temporal history are retained.

Let S^I_{ij} be a symmetric trace-free intrinsic shape tensor in the local orthonormal rest frame of the galaxy population. As in Chapter 1, average over short-scale formation physics while holding fixed the long-wavelength tidal history. Define the linear response kernel

$$\mathcal{R}_{ij}{}^{kl}(\tau;\tau') \equiv \left.\frac{\delta \left\langle S^I_{ij}(\tau)\right\rangle_{S\,|\,\mathcal{E}_L}}{\delta \mathcal{E}_{kl}(\tau')}\right|_{\mathcal{E}_L=0}. \tag{5.9}$$

Spatial locality places the perturbation on the past worldline; time nonlocality is retained through τ'. In an isotropic unperturbed ensemble, the only rank-four tensor mapping one symmetric trace-free tensor into another is the STF projector

$$\mathcal{P}_{ij}{}^{kl} \equiv \frac{1}{2}(\delta_i{}^k\delta_j{}^l + \delta_i{}^l\delta_j{}^k) - \frac{1}{3}\delta_{ij}\delta^{kl}. \tag{5.10}$$

Therefore

$$\mathcal{R}_{ij}{}^{kl}(\tau;\tau') = R_E(\tau;\tau')\mathcal{P}_{ij}{}^{kl}, \tag{5.11}$$

and the leading intrinsic response is

$$\boxed{S^I_{ij}(\tau) = \int^{\tau} \mathrm{d}\tau'\, R_E(\tau;\tau')\mathcal{E}_{\langle ij\rangle}(\mathbf{x}_{\mathrm{fl}}(\tau'),\tau') + \epsilon^I_{ij} + \cdots}. \tag{5.12}$$

Rotational invariance fixes the index structure but not the temporal kernel R_E. The latter contains the internal dynamics of the selected population: a rapidly adjusting system, a fossil response set at formation, and a population with a broad distribution of formation times can all have different kernels while obeying the same tensor symmetry. This is the tensor counterpart of the time nonlocality found for scalar bias in Eq. 1.13, and it carries the same warning. A measurement of the alignment amplitude at one redshift constrains an integral of R_E, not its instantaneous value.

The stochastic shape ϵ^I_{ij} has zero conditional mean. Reorganizing the time integral perturbatively gives the familiar local-at-observation-time expansion

$$S^I_{ij} = b_K K_{ij} + b_{\delta K}\delta K_{ij} + b_{KK}\left(K_{ik}K^k{}_j - \frac{1}{3}\delta_{ij}K_{kl}K^{kl}\right) + \cdots. \tag{5.13}$$

The normalization of b_K depends on whether the operator is written as K_{ij} or as the dimensionful curvature $\mathcal{E}_{ij}$. The response definition Eq. 5.9 fixes that convention.

The observed intrinsic shear is the screen-projected trace-free tensor

$$\gamma^I_{ab} = \left(P_a{}^iP_b{}^j - \frac{1}{2}P_{ab}P^{ij}\right)S^I_{ij}. \tag{5.14}$$

Projection converts the local three-dimensional tensor into a spin-two field on the sky; it does not make the source response nonlocal.

The screen projection also removes the component along the line of sight and the two-dimensional trace. It is a kinematic step performed after the local response has been defined. Correlations of γ^I_{ab} with density or lensing fields then follow from correlations of the long-wavelength tide with those observables.

5.3 A gravitational wave as a local tidal field

The transverse-traceless metric is a convenient coordinate representation of a gravitational wave, but the local detector responds to curvature. Computing R_{0i0j} makes this statement explicit and permits a direct comparison with the scalar tidal field. The result is independent of whether the same wave is described in TT coordinates or in Fermi coordinates around the detector.

In a local region small compared with the background curvature radius, take

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + [\delta_{ij} + h^{\mathrm{TT}}_{ij}(t, \mathbf{x})]\mathrm{d}x^i\mathrm{d}x^j, \tag{5.15}$$

with

$$\partial^i h^{\mathrm{TT}}_{ij} = 0, \qquad h^{\mathrm{TT}\,i}{}_i = 0, \qquad h_{0\mu} = 0. \tag{5.16}$$

Substituting these conditions into the general linearized curvature Eq. 5.2,

$$\begin{aligned} R^{\mathrm{GW}}_{0i0j} &= \frac{1}{2}\left(h_{0j,i0} + h_{i0,0j} - h_{00,ij} - h_{ij,00}\right) \\ &= -\frac{1}{2}\ddot{h}^{\mathrm{TT}}_{ij}. \end{aligned} \tag{5.17}$$

The local equation of geodesic deviation is therefore

$$\boxed{\ddot{X}^i = \frac{1}{2}\ddot{h}^i{}_j{}^{\mathrm{TT}} X^j}. \tag{5.18}$$

The corresponding Fermi metric begins with

$$g^{\mathrm{F}}_{\hat{0}\hat{0}} = -1 + \frac{1}{2}\ddot{h}^{\mathrm{TT}}_{ij}(T, \mathbf{0})X^iX^j + \mathcal{O}(X^3). \tag{5.19}$$

A spatially constant value of h^{TT}_{ij} does not appear in the local metric; the observable source-side effect begins with the curvature, or two derivatives of the tensor perturbation.

For a wave of frequency ω, the local tidal amplitude therefore scales as $\omega^2 h^{\mathrm{TT}}_{ij}$. A detector or a galaxy responds after integrating this tidal acceleration against its own dynamical response. The final displacement can be proportional to h^{TT}_{ij} for freely falling test masses because the equation of motion is integrated twice in time; the local forcing itself remains the curvature.

5.4 The ring experiment

The ring of freely falling particles separates the coordinate description from the measured effect. The central particle defines the Fermi frame, and the other particles supply deviation vectors. The polarization tensors determine the eigenvectors and eigenvalues of the tidal matrix in the plane transverse to propagation.

Figure placeholder. The ring experiment. A ring of freely falling test particles in the plane transverse to a wave propagating along z, shown at four phases of the cycle for the $+$ polarization and for the $\times$ polarization, with the central particle marked as the origin of the Fermi frame and the principal axes of the two cases drawn to show the 45° offset between them.

For a wave propagating in the z direction,

$$h_{ij}^{\mathrm{TT}} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} . \tag{5.20}$$

Eq. 5.18 gives

$$\begin{aligned} \ddot{X} &= \frac{1}{2}\ddot{h}_+ X + \frac{1}{2}\ddot{h}_\times Y, \\ \ddot{Y} &= \frac{1}{2}\ddot{h}_\times X - \frac{1}{2}\ddot{h}_+ Y. \end{aligned} \tag{5.21}$$

To first order, replace (X, Y) on the right-hand side by the unperturbed position (X_0, Y_0). If h_{ij} and $\dot{h}_{ij}$ vanish at the reference time and the particles are initially at rest relative to the central geodesic, two integrations give

$$X^i(T) = \left(\delta^i{}_j + \frac{1}{2}h^i{}_j{}^{\mathrm{TT}}(T)\right) X_0^j. \tag{5.22}$$

A circle becomes an ellipse whose principal axes rotate by 45° between the $+$ and $\times$ polarizations. In TT coordinates the particles may remain at fixed coordinates while their proper separation changes; in FNC the central observer describes the same physics as relative acceleration. The common invariant is Eq. 5.17.

The initial conditions in Eq. 5.22 remove homogeneous relative position and velocity modes that are unrelated to the passing wave. Other initial conditions add the corresponding free relative motion. The wave-induced part is still obtained from the same geodesic-deviation equation.

5.5 Intrinsic response to a tensor tide

Because the local response is organized by tensor representation and curvature, the source need not know whether a trace-free tide originated from scalar density perturbations or from a propagating tensor mode. The distinction enters through the spacetime and time dependence of $\mathcal{E}_{ij}$ and through the correlations of that tide with other observables.

The response kernel Eq. 5.9 does not distinguish whether the trace-free curvature was generated by scalar or tensor perturbations. For a gravitational wave,

$$\mathcal{E}^{\mathrm{GW}}_{ij} = R^{\mathrm{GW}}_{\hat{0}i\hat{0}j} = -\frac{1}{2}\ddot{h}^{\mathrm{TT}}_{ij}. \tag{5.23}$$

Hence

$$S^{I,\mathrm{GW}}_{ij}(\tau) = -\frac{1}{2}\int^{\tau} \mathrm{d}\tau'\, R_E(\tau;\tau')\ddot{h}^{\mathrm{TT}}_{ij}(\tau') + \cdots. \tag{5.24}$$

For a monochromatic local wave this response carries the expected factor ω^2 before the temporal response kernel is evaluated. The suppression is an equivalence-principle statement: a local source does not respond to a constant tensor metric perturbation, but it can respond to the associated tidal history.

Whether the final intrinsic shape is additionally suppressed or enhanced depends on the ratio of the wave frequency to the internal dynamical and formation timescales represented by R_E. The curvature factor is universal; the temporal transfer function is population dependent.

5.6 Intrinsic source shape versus propagated ruler shear

The same observed spin-two field can contain a local source response and a propagation distortion. Their tensor appearance alone does not identify the mechanism. The distinction is made by where the response is defined: intrinsic alignment is a functional of local tidal history on the source worldline, whereas ruler shear is a functional of the null geodesic and its endpoints.

Figure placeholder. Intrinsic alignment against ruler shear. A single spacetime diagram carrying both mechanisms: on the source worldline, the local tidal history that sets S^I_{ij}, drawn as a tetrad with a distorted galaxy shape; along the null geodesic reaching the observer, the accumulated deflection that produces $\gamma^{\mathrm{ruler}}_{ab}$. The point of the figure is that both terminate in the same measured ellipticity on the observer's screen.

The intrinsic tensor Eq. 5.14 is defined at the source. The ruler shear of Chapter 3 is generated by the map from source to observer:

$$\gamma^{\mathrm{ruler}}_{ab} = -\frac{1}{2}\left(P_a{}^i P_b{}^j - \frac{1}{2}P_{ab}P^{ij}\right) h_{ij}\big|_s - \partial_{\perp\langle a}\Delta x_{\perp b\rangle}. \tag{5.25}$$

The displacement contains line-of-sight integrals and observer terms. Schematically, an observed shape estimator receives

$$\gamma^{\mathrm{obs}}_{ab} = \gamma^I_{ab} + \gamma^{\mathrm{ruler}}_{ab} + \text{selection and calibration responses}. \tag{5.26}$$

Both terms are spin two and can be correlated with the same long mode, but they encode different mechanisms:

$$\text{local formation response} \neq \text{null-propagation distortion}. \tag{5.27}$$

FNC isolates the first. Cosmic rulers organize the second. A consistent observable prediction requires both.

This separation is also useful for gauge reasoning. The intrinsic tensor is defined in a local orthonormal tetrad and is invariant under a change of cosmological coordinates. The propagated contribution is gauge-invariant only after source, line-of-sight, and observer terms are combined. Their sum can then be projected into whatever shape estimator and selection convention the survey uses.

> **! Key idea: the local response sees a tensor, not a mechanism**
>
> The response kernel Eq. 5.9 is blind to the origin of the tide it integrates. A trace-free curvature produced by the collapse of a long-wavelength density mode and one produced by a passing gravitational wave enter the same way, with the same universal factor and the same population-dependent kernel R_E. This is a direct consequence of the local-frame argument of Chapter 4: what reaches the source is $\mathcal{E}_{\langle ij \rangle}$ and nothing else. Two practical statements follow. Physical mechanisms are distinguished by the time dependence of the tide and by its correlation with other observables, never by the response coefficient alone. And an observed spin-two field is not automatically a source property, since Eq. 5.26 shows it also contains a propagation term that is a functional of the null geodesic rather than of the source worldline.

Conclusion: from spacetime fields to measured structure

These lectures began with a Newtonian statement, $\delta_g = b_1 \delta_m$, and asked what must be added before it becomes a prediction for a relativistic observation. The answer was not a catalogue of small corrections. It was a sequence of maps with different physical roles.

The bias expansion describes how a local population responds to its physical environment. The null-geodesic initial-value problem maps the energy and direction measured by an observer to a source event. The cosmic clock compares the proper age of that event with the age assigned by its observed redshift. Cosmic rulers compare physical separations in the source rest frame with separations inferred from angles and redshifts. Fermi normal coordinates then identify the curvature components that can affect local source physics after the equivalence principle has removed a uniform gravitational field.

The same organizing principle runs through all five lectures:

$$\boxed{\text{define the measurement covariantly} \longrightarrow \text{choose a perturbative representation} \longrightarrow \text{combine source, propaga}}$$

Diffeomorphism invariance is therefore not an abstract redundancy that disappears once a gauge has been chosen. It constrains how a fundamental theory becomes a prediction for data. A proposed explanation of the dark universe must pass through this chain: it must determine the local degrees of freedom and their dynamics, their effect on tracers and light propagation, and finally the gauge-independent relations among observed redshifts, directions, shapes, and counts.

Large-scale structure is demanding precisely because it requires all of these ingredients simultaneously. It is also valuable for the same reason. New gravitational fields, modified interactions, nonstandard dark matter, or early-universe initial conditions can be tested only after their spacetime dynamics have been translated into the operational observables constructed here. The framework does not select the new idea. It supplies the standard that a new idea must meet before it can be compared consistently with the observed universe.

Summary

Key ideas. The electric Riemann tensor in the observer's tetrad is the single object behind scalar tidal bias, intrinsic alignment, and the local effect of a gravitational wave. A scalar abundance cannot respond linearly to a trace-free tide, while a shape tensor can, which is why intrinsic alignment appears at linear order in K_{ij} whereas scalar clustering feels the tide first at quadratic order. Rotational invariance fixes the index structure of the shape response completely, leaving one scalar kernel that encodes the internal dynamics of the population and remains nonlocal in time. Finally, a local source responds to a tensor representation rather than to a mechanism, so scalar and tensor tides enter identically and are distinguished only by their time dependence and correlations.

Main results.

- The scalar tide as a curvature component, Eq. 5.7, and its identification with the bias operator, Eq. 5.8, closing the argument opened at Eq. 1.18.
- The tensor-valued response kernel Eq. 5.9, its isotropic reduction Eq. 5.11, and the intrinsic shape Eq. 5.12 with its local expansion Eq. 5.13.
- The screen projection Eq. 5.14 converting a local three-tensor into a spin-two field on the sky.
- The gravitational-wave curvature Eq. 5.17, the geodesic deviation Eq. 5.18, and the ring solution Eq. 5.22.
- The intrinsic response to a tensor tide Eq. 5.24, carrying the equivalence-principle factor ω^2.
- The separation of intrinsic and propagated shear, Eq. 5.26.

Where this leaves these notes. The two notions of observability introduced in the Preface have now both been constructed. Part I produced a nonlocal observable on the past light cone, Eq. 3.38, in which source, propagation, and observer terms combine into a gauge-invariant count. Part II produced local observables in a freely falling frame, where the connection is removed and curvature remains. A complete prediction for a survey requires both, and the last section of this chapter shows why: the same measured spin-two field on the sky contains a local formation response organized by Chapter 4 and a propagation distortion organized by Chapter 3.

References

1. Desjacques, V., Jeong, D. & Schmidt, F. Large-scale galaxy bias. *Physics Reports* **733**, 1–193 (2018).
2. Schmidt, F. & Jeong, D. Cosmic rulers. *Physical Review D* **86**, 083527 (2012).
3. Jeong, D. & Schmidt, F. Cosmic clocks. *Physical Review D* **89**, 043519 (2014).
4. Ginat, Y. B., Desjacques, V., Jeong, D. & Schmidt, F. Covariant decomposition of the non-linear galaxy number counts and their monopole. *Journal of Cosmology and Astroparticle Physics* **12**, 031 (2021).
5. Manasse, F. K. & Misner, C. W. Fermi normal coordinates and some basic concepts in differential geometry. *Journal of Mathematical Physics* **4**, 735–745 (1963).
6. Kaiser, N. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society* **227**, 1–21 (1987).
7. Bardeen, J. M. Gauge-invariant cosmological perturbations. *Physical Review D* **22**, 1882–1905 (1980).
8. Ellis, G. F. R. Relativistic cosmology. in *General relativity and cosmology* (ed. Sachs, R. K.) vol. 47 104–182 (Academic Press, New York, 1971).
9. Fermi, E. Sopra i fenomeni che avvengono in vicinanza di una linea oraria. *Atti della Reale Accademia dei Lincei, Rendiconti, Serie 5* **31**, 21–23, 51–52, 101–103 (1922).

A Response conventions and functional derivatives

This appendix collects normalization choices that are easy to conflate in applications. A derivative of the abundance and a derivative of its logarithm agree at first order but differ at nonlinear order, and a functional response kernel should be distinguished from the coefficients of a chosen local operator basis.

The response coefficients can be defined using derivatives of either $\bar{n}_g$ or $\ln \bar{n}_g$. For a homogeneous density background Δ, define

$$b_N = \frac{1}{\bar{n}_g} \frac{\partial^N \bar{n}_g}{\partial \Delta^N}\bigg|_0, \qquad \beta_N = \frac{\partial^N \ln \bar{n}_g}{\partial \Delta^N}\bigg|_0. \tag{A.1}$$

Then

$$b_1 = \beta_1, \qquad b_2 = \beta_2 + \beta_1^2, \qquad b_3 = \beta_3 + 3\beta_1\beta_2 + \beta_1^3. \tag{A.2}$$

The notes use the b_N convention because it is the coefficient in the Taylor expansion of the fractional abundance.

For a functional $F[\mathcal{O}]$, the derivative is defined by

$$\delta F = \int \mathrm{d}^4x \, \frac{\delta F}{\delta \mathcal{O}^A(x)} \delta \mathcal{O}^A(x). \tag{A.3}$$

The response kernel in Eq. 1.9 is a distribution. Spatial locality does not mean it is exactly proportional to a spatial Dirac delta. Rather, the kernel has support over a region of size R_* and admits the derivative expansion

$$\int \mathrm{d}^3r \, W_R(\mathbf{r}) \mathcal{O}(\mathbf{x} + \mathbf{r}) = \mathcal{O}(\mathbf{x}) + c_2 R_*^2 \nabla^2 \mathcal{O}(\mathbf{x}) + \cdots \tag{A.4}$$

for long modes. Eq. 1.12 keeps the leading term.

A scalar separate-universe response probes only isotropic backgrounds. Tidal response requires an anisotropic long-wavelength background, and tensor-valued response such as Eq. 5.9 requires retaining the free tensor indices before imposing rotational invariance.

B Pullback and Lie-derivative identities

The differential-form construction in Chapter 2 uses only the standard functorial properties of pullbacks and Cartan's formula. They are listed here to make clear which steps are geometric identities and which steps use cosmological dynamics.

For a smooth map $F : N \to M$ and a p-form α on M, the pullback is defined by

$$(F^*\alpha)_q(v_1, \dots, v_p) = \alpha_{F(q)}(F_*v_1, \dots, F_*v_p). \tag{B.1}$$

It satisfies

$$\begin{aligned} F^*(\alpha \wedge \beta) &= F^*\alpha \wedge F^*\beta, \\ F^*(\mathrm{d}\alpha) &= \mathrm{d}(F^*\alpha), \\ (F \circ G)^* &= G^*F^*. \end{aligned} \tag{B.2}$$

For a vector field X with flow φ_λ,

$$\mathcal{L}_X\alpha = \frac{\mathrm{d}}{\mathrm{d}\lambda}\varphi_\lambda^*\alpha\Big|_{\lambda=0} = \mathrm{d}(\iota_X\alpha) + \iota_X\mathrm{d}\alpha. \tag{B.3}$$

These identities are sufficient to derive both current conservation and the linear pullback formula.

Two nowhere-vanishing top forms on the same oriented three-manifold differ by a scalar function. Thus the definition

$$\mathsf{X}^*\mathcal{J} = n_g^{\mathrm{obs}}\widetilde{\boldsymbol{\omega}} \tag{B.4}$$

is coordinate-independent. The scalar n_g^{obs} is the Radon–Nikodym-type ratio of the physical count form to the fiducial observed-coordinate volume form.

C Gauge-transformation table

The table uses one passive convention throughout. When comparing formulas across references, the sign of the gauge generator and the sign convention for the scalar spatial metric perturbation should be translated before individual terms are compared.

A tilde is reserved for observed or background-inferred quantities. For the passive relabelling

$$x^\mu \mapsto x^\mu + (T, \partial^i L),$$

the conventions used in these notes give

Table C.1: Gauge transformations in the passive convention of Eq. 2.1.

Quantity	Transformation
A	$A \mapsto A - \mathcal{H}T - T'$
B	$B \mapsto B + L' - T$
D	$D \mapsto D - \mathcal{H}T$
E	$E \mapsto E - L$
v	$v \mapsto v + L'$
δ_m	$\delta_m \mapsto \delta_m + 3\mathcal{H}T$
δ_g	$\delta_g \mapsto \delta_g - (\bar{n}_g'/\bar{n}_g)T$
Δx^μ	$\Delta x^\mu \mapsto \Delta x^\mu + \xi^\mu$
$\Delta \ln a$	$\Delta \ln a \mapsto \Delta \ln a + \mathcal{H}T$
I_A	$I_A \mapsto I_A - aT$
$\mathcal{T} = HI_A + \Delta \ln a$	invariant
$\delta_m^{\rm pt} = \delta_m + 3HI_A$	invariant
$\delta_m^{\rm or} = \delta_m - 3\Delta \ln a$	invariant
$\delta\mathcal{J} + \mathcal{L}_{\Delta x}\overline{\mathcal{J}}$	invariant
$\mathcal{C}, \mathcal{B}_i, \mathcal{A}_{ij}$	invariant as complete ruler distortions

For any scalar background $\bar{S}(\eta)$,

$$\delta S \mapsto \delta S - \bar{S}'T. \tag{C.1}$$

The signs of the density transformations are fixed by the background evolution. The table uses the boundary convention $(aT)_{\rm ini} = 0$. Without it, the I_A row reads $I_A \mapsto I_A - aT + (aT)_{\rm ini}$; see Sec. 2.5.

C.1 Gauge-invariant combinations

The gauge transformations listed in Table C.1 immediately identify the scalar shear variable

$$\sigma \equiv B + E'.$$

Using

$$B \mapsto B + L' - T, \qquad E \mapsto E - L,$$

one finds

$$\sigma \mapsto (B + L' - T) + (E' - L') = \sigma - T.$$

The cancellation of the spatial gauge generator L is automatic, so only the time shift T remains.

The two Bardeen potentials adapted to the metric convention adopted in these notes are therefore

$$\boxed{\Psi_B = A - \mathcal{H}\sigma - \sigma'}$$

and

$$\boxed{\Phi_B = D - \mathcal{H}\sigma.}$$

Substituting the transformation rules from Table C.1 shows that both combinations are invariant under an arbitrary scalar gauge transformation.

In the Newtonian (longitudinal) gauge,

$$B = E = 0,$$

so that

$$\Psi_B = A, \qquad \Phi_B = D.$$

C.1.1 Remark on conventions

Many cosmology references instead write the spatial metric as

$$g_{ij} = a^2(1 - 2\Phi)\delta_{ij},$$

whereas these notes use

$$g_{ij} = a^2(1 + 2D)\delta_{ij}.$$

Accordingly,

$$D = -\Phi_{\text{common}},$$

where $\Phi_{\rm common}$ denotes the potential in the more common convention.

More generally, the definition of the scalar shear variable depends on the complete metric convention (including the sign of g_{0i}, the definition of E, and the passive/active gauge convention). Gauge-transformation formulae should therefore always be translated into a common convention before comparing expressions from different references.

D Manasse–Misner index order and the quadratic algebra

The original Manasse–Misner paper uses an index placement that can make its printed signs appear opposite to modern all-lowered conventions. The following translation keeps the mixed tensor fixed and shows where the sign enters when the first index pair is lowered and reordered.

Manasse and Misner write

$$R_{\mu}{}^{\nu}{}_{\alpha\beta}, \tag{D.1}$$

whereas these notes write the same mixed tensor as

$$R^{\nu}{}_{\mu\alpha\beta}. \tag{D.2}$$

After lowering the contravariant index, their ordering gives

$$R^{\mathrm{MM}}_{\mu\nu\alpha\beta} = g_{\nu\rho} R^{\rho}{}_{\mu\alpha\beta} = R_{\nu\mu\alpha\beta} = -R_{\mu\nu\alpha\beta}. \tag{D.3}$$

Thus their printed expansion

$$\begin{aligned} g_{00} &= -1 + R^{\mathrm{MM}}_{0i0j} X^i X^j + \cdots, \\ g_{0i} &= \frac{2}{3} R^{\mathrm{MM}}_{0jik} X^j X^k + \cdots, \\ g_{ij} &= \delta_{ij} + \frac{1}{3} R^{\mathrm{MM}}_{ikjl} X^k X^l + \cdots \end{aligned} \tag{D.4}$$

becomes Eq. 4.48, Eq. 4.49, and Eq. 4.50.

For completeness, the mixed component algebra in Eq. 4.46 is

$$\begin{aligned} g_{\hat{0}k,ij} &= -\Gamma^{\hat{0}}_{ki,j} + \Gamma^{k}_{\hat{0}i,j} \\ &= -\frac{2}{3}(R_{\hat{0}ikj} + R_{\hat{0}jki}), \end{aligned} \tag{D.5}$$

where Eq. 4.39 and Eq. 4.42, pair exchange, and the first Bianchi identity have been used. Likewise,

$$\begin{aligned} g_{\ell m,ij} &= \Gamma^{\ell}_{mi,j} + \Gamma^{m}_{\ell i,j} \\ &= -\frac{1}{3}(R_{\ell imj} + R_{\ell jmi}). \end{aligned} \tag{D.6}$$

Because $X^i X^j$ is symmetric, these second-derivative expressions reduce to the compact metric coefficients in the main text.