Causal Information Loss and the Inverse Problem: A Theory of Latent Information, Causal Revelation, and Reversibility

Kweku Opoku-Agyemang*

September 2026

Abstract

Economic reasoning is usually directional: we ask what an intervention does to an outcome. But economists also routinely face the reverse problem—given an observed outcome, what can be learned about the causal state that produced it? This paper develops a theory of when such causal reversal is possible, what information a forward causal mechanism destroys, and how much additional information is required for reconstruction. For a structural causal mechanism, we associate each latent state with its complete response to interventions on the causal antecedent. This response-function map induces a canonical causal quotient: latent states that generate the same response to every intervention are identified. We show that this quotient is the minimal latent object relevant for structural inversion. Exact inverse recovery has two distinct sources of difficulty: the mechanism may be structurally noninvertible, or the causally relevant latent information may be unavailable. We then characterize how much of this latent ambiguity is revealed by the observed outcome. Under regularity conditions, the information revealed by the outcome equals the reduction in the intrinsic dimension of the unresolved causal state. Finally, we define an inverse-causal rate-distortion function that measures the minimum additional information required to reconstruct the causal antecedent at a given resolution. Its high-resolution behavior is governed by the dimension of the unresolved causal state. The resulting theory provides a unified characterization of causal reversibility, causal information loss, and inverse information requirements, linking structural causal models, information theory, differential geometry, and inverse problems.

*Machine Learning X Doing Incorporated. Email: kweku@machinelearningxdoing.com.

1 Introduction

Causal inference is fundamentally directional. A structural causal model specifies how an antecedent, together with latent states of the world, generates an effect. Given a causal mechanism

$$Y = f(X, U),$$

the conventional problem is to determine the effect of interventions on X, or to identify features of the distribution of Y under such interventions.

The inverse problem asks a different question: given the realized effect Y, to what extent can the causal antecedent X be recovered? More fundamentally, when is a causal mechanism reversible, what information is lost when it operates in the forward direction, and what additional information is required to reverse it?

This distinction is important because a question about the *effect of a cause* is not equivalent to a question about the *cause of an effect*. In the forward direction, the causal variable of interest can be specified and its intervention compared with a suitable counterfactual. In the reverse direction, however, observing a realized effect does not in general reveal which combination of causal states and mechanisms produced it. For an individual realization, the counterfactual states needed to uniquely attribute the observed outcome are not jointly observed. Thus, within a statistical framework, reverse questions about the causes of a particular effect are generally not direct counterparts of conventional causal-effect estimation.

This does not make reverse causal questions meaningless. They arise naturally whenever an observed outcome is anomalous, when a historical event demands explanation, or when an investigator seeks to reconstruct the state of a system from its consequences. Such questions can motivate hypotheses and guide model construction. The difficulty is instead to characterize precisely what can be recovered from the effect, and what additional structure or information would be required to make the reverse problem well-defined.

The present paper takes a structural approach to this question. Rather than attempting to infer an arbitrary set of causes from a realized outcome, we take a causal mechanism as given and ask:

What information is required to reverse that mechanism?

This perspective leads to a distinction between two fundamentally different obstacles to inverse recovery. The causal mechanism itself may be structurally non-invertible: even if the latent state were completely known, different antecedents may generate the same effect. Alternatively, the mechanism may be structurally invertible conditional on its latent state, while the information needed to identify that state is unavailable. The purpose of this paper is to characterize this second problem, its interaction with structural non-invertibility, and the information requirements of reversal.

These questions are not answered by Bayes' rule alone. The observational conditional distribution

$$P(X \mid Y)$$

describes uncertainty about the antecedent after observing the effect, but it does not in general constitute an inverse causal operator. In a noisy causal mechanism, different antecedents and latent states can generate the same effect. Knowledge of the distribution of the latent disturbance does not reveal its realized value. Thus, even when the forward causal mechanism is completely known, the realized causal antecedent need not be identifiable from the effect.

This paper develops a general theory of this phenomenon. Our central object is the information about the latent state that remains relevant to the causal response function. For every latent state u, define

$$f_u(x) = f(x, u),$$

and introduce the response-function map

$$\Phi_f : \mathcal{U} \longrightarrow \mathcal{F}(\mathcal{X}, \mathcal{Y}), \qquad \Phi_f(u) = f_u.$$

Two latent states are causally equivalent when they induce the same response function:

$$u \sim_f u' \quad \Longleftrightarrow \quad f(x, u) = f(x, u') \quad \text{for all } x.$$

The resulting quotient

$$Q = \Phi_f(U)$$

is the minimal latent information that is relevant for distinguishing causal response functions. Importantly, this is a causal rather than merely observational equivalence relation: two latent states are identified only when they are indistinguishable under the entire family of relevant interventions on X, not merely under the realized or observational distribution of X.

This response-function quotient provides a natural foundation for a theory of causal information loss. The forward mechanism

$$U \longrightarrow Y$$

may erase distinctions among latent states that matter neither for the observed effect nor for any counterfactual response. Such distinctions should not be counted as information that must be recovered in order to reverse the mechanism. The relevant information loss is instead the loss induced by the map

$$U \xrightarrow{\Phi_f} Q.$$

For discrete U, this loss can be represented by

$$H(U \mid Q).$$

In continuous systems, however, entropy is not the appropriate primitive object. The corresponding structural object is the sigma-field

$$\sigma(Q) = \sigma(\Phi_f(U)) \subseteq \sigma(U),$$

which records precisely which latent distinctions remain causally relevant.

The quotient also clarifies the conditions for inverse identification. Suppose that, conditional on the quotient, the structural mechanism can be written as

$$Y = f_Q(X)$$

and that f_Q is invertible in X. Then

$$X = f_Q^{-1}(Y).$$

Thus exact structural reversal requires two conceptually distinct conditions. First, the mechanism must be structurally invertible in the antecedent. Second, the causally relevant quotient Q must be available. Failure of the first condition is an intrinsic ambiguity of the mechanism: even complete knowledge of the latent state cannot distinguish some antecedents. Failure of the second is an information deficiency: the mechanism is reversible conditional on the causal state, but that state is not sufficiently observed.

This decomposition is particularly transparent in additive-noise models. Consider

$$Y = h(X) + U,$$

where h is injective. Conditional on U,

$$X = h^{-1}(Y - U),$$

so the mechanism is structurally invertible. But if U is unobserved, the realized antecedent generally cannot be recovered exactly from Y. In this model,

$$\Phi_f(u)(x) = h(x) + u,$$

and the response-function map is injective: distinct values of U correspond to distinct causal response functions. Hence the entire latent disturbance, rather than merely a lower-dimensional summary, is relevant for structural inversion. By contrast, in the model

$$Y = X + U^2,$$

the response function is

$$\Phi_f(u)(x) = x + u^2,$$

so u and $-u$ are causally equivalent. The minimal quotient is $Q = U^2$, and the sign of U is causally redundant for inverse recovery. The quotient construction therefore distinguishes information that is genuinely required for reversal from information that is merely present in the latent state.

The next question is how much of this latent causal ambiguity is already revealed by the observed effect. Let

$$Y = F(X, Q)$$

be a smooth representation of the mechanism in terms of the causal quotient. For a realized effect y, define the observation fiber

$$\mathcal{F}_y = \{(x, q) : F(x, q) = y\},$$

and its projection onto quotient space,

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y) = \{q : \exists x,\ F(x, q) = y\}.$$

The latter, rather than the full fiber $\mathcal{F}_y$, is the relevant geometric object for inverse identification: X is the unknown antecedent and therefore a nuisance coordinate, whereas Q indexes the latent causal response that must be distinguished.

We characterize the local dimension of this projected fiber by removing from the effect space all directions that can be generated by changing X. At (x, q), let

$$\mathcal{S}_X = \operatorname{Im} D_x F$$

and define the induced quotient derivative

$$\overline{D}_q F : T_q\mathcal{Q} \longrightarrow T_y\mathcal{Y}/\mathcal{S}_X.$$

Its rank,

$$r_{\text{rev}} = \operatorname{rank}(\overline{D}_q F),$$

is the *causal revelation rank*. It measures the number of locally independent directions in the causal quotient whose effects cannot be replicated by perturbing the unknown antecedent.

Our first main geometric result shows that, under constant-rank and global projection regularity conditions, the projected observation fiber is an embedded submanifold of $\mathcal{Q}$ with dimension

$$\dim \mathcal{A}_y = \dim \mathcal{Q} - r_{\mathrm{rev}}.$$

The distinction between immersion and embedding is important here. Constant-rank arguments provide the local manifold structure, while properness and injectivity of the restricted projection ensure that the resulting image is an embedded submanifold rather than a self-intersecting or otherwise nonembedded immersed manifold.

The geometric result has a direct probabilistic interpretation. Let

$$P_{Q|Y=y}$$

denote a regular conditional distribution of the causal quotient given the observed effect. Under a regularity condition ensuring that this conditional law is absolutely continuous with respect to the natural Hausdorff measure on $\mathcal{A}_y$, with a positive finite density almost everywhere, its information dimension equals the dimension of the projected fiber. Consequently,

$$d_I(Q \mid Y = y) = \dim \mathcal{A}_y = k - r_{\mathrm{rev}},$$

where $k = \dim \mathcal{Q}$. When Q itself has information dimension k, this yields the causal revelation identity

$$\boxed{d_I(Q) - d_I(Q \mid Y) = r_{\mathrm{rev}}.}$$

The differential rank therefore has a precise information-theoretic interpretation: it is the dimensional amount of causal ambiguity revealed by the effect.

This result also clarifies an important distinction between causal relevance and observational revelation. The response-function map Φ_f determines which latent distinctions are causally meaningful. Conditional on the realized effect, the law $P(Q \mid Y)$ determines which of those distinctions remain unresolved. Thus

$$\Phi_f \quad \text{defines causal relevance},$$

whereas

$$P(Q \mid Y) \quad \text{measures observational revelation}.$$

The two operations should not be conflated.

The framework yields a natural quantitative theory of the information required for inverse recovery. For discrete variables, if exact structural inversion is possible conditional on Q, then

$$H(X \mid Y) = I(X; Q \mid Y).$$

The right-hand side measures the additional causal information required after observing the effect. For continuous variables, exact inversion can make conditional mutual information infinite, and differential entropy is not invariant enough to serve as a universal primitive. We therefore define an operational inverse-causal rate-distortion function,

$$R_{\mathrm{IC}}(\varepsilon) = \inf I(Q; M \mid Y),$$

where the infimum is taken over representations M of the causal quotient and decoders producing $\widehat{X}$ satisfying

$$\mathbb{E}[d_X(X, \widehat{X})] \le \varepsilon.$$

This quantity measures the minimum information about the causally relevant latent state required to recover the antecedent to resolution ε.

Under regularity conditions linking the inverse map to the geometry of the quotient, we show that the high-resolution behavior of this operational quantity is governed by conditional information dimension:

$$R_{\mathrm{IC}}(\varepsilon) = \frac{1}{2} d_I(Q \mid Y) \log \frac{1}{\varepsilon} + o\left(\log \frac{1}{\varepsilon}\right).$$

Combining this result with the causal revelation identity gives

$$R_{\mathrm{IC}}(\varepsilon) = \frac{1}{2}\left[d_I(Q) - r_{\mathrm{rev}}\right] \log \frac{1}{\varepsilon} + o\left(\log \frac{1}{\varepsilon}\right).$$

Thus inverse difficulty is determined by the causal dimensional complexity of the latent response state after subtracting the dimensions already revealed by the effect.

The theory consequently separates four phenomena that are often conflated in inverse problems. First, a causal mechanism may be *structurally non-invertible*, in which case even complete knowledge of the latent state cannot recover the antecedent. Second, it may be structurally invertible but *information-deficient*, because the causal quotient is not observed. Third, the latent state may contain *causally redundant information*, which is removed by the response-function quotient. Fourth, the observed effect may reveal some, but not all, of the remaining quotient through the geometry of the observation fibers. These distinctions yield the conceptual decomposition

$$\boxed{\text{inverse difficulty} = \text{causal latent complexity} - \text{causal revelation.}}$$

The paper contributes to several strands of the literature (summarized in the next section). Relative to causal inference, it develops a formal theory of the inverse direction of a structural causal mechanism rather than treating $P(X \mid Y)$ as an implicit inverse. Relative to econometric identification, it separates structural ambiguity from missing latent information and provides a geometric characterization of the information that must be supplied for reversal. Relative to information theory, it identifies a causal quotient as the appropriate object whose conditional information dimension governs inverse reconstruction, rather than charging for latent distinctions that have no causal effect on the response function. Relative to differential geometry, it uses projected observation fibers and quotient tangent maps to characterize the dimensions of unresolved causal states.

The remainder of the paper develops these results formally. Section **??** introduces the structural framework and the response-function quotient. Section **??** establishes the decomposition of inverse identification into structural invertibility and quotient availability. Section **??** develops the projected-fiber geometry and proves the causal revelation theorem. Section **??** connects the geometric result to conditional information dimension. Section **??** develops the operational inverse-causal rate-distortion formulation and its high-resolution asymptotics. Section **??** studies canonical examples, including additive-noise and nonlinear latent mechanisms. Section **??** discusses implications for causal inference, econometrics, and inverse problems.

2 Related Literature

Our analysis lies at the intersection of four literatures: causal identification and structural causal models, inverse problems and latent-variable identification, information-theoretic measures of dependence and dimension, and geometric analysis of fibers and conditional distributions. The contribution is not to introduce a new method within any one of these literatures in isolation. Rather, we develop a common object—the causal quotient induced by a structural response-function map—and show that it connects structural reversibility, geometric causal revelation, information dimension, and inverse-causal rate distortion.

2.1 Causal identification and structural mechanisms

The modern causal-inference literature provides several complementary languages for reasoning about causal mechanisms and identification. The potential-outcomes framework emphasizes counterfactual responses to interventions, while structural causal models represent these responses through functions of treatments and latent disturbances. The econometric literature has developed identification theory for a wide range of structural and semiparametric models, with particular emphasis on the conditions under which causal or structural objects are uniquely determined by observable distributions. See, among others, Manski [1990], Pearl [2009], Imbens and Rubin [2015], Hernán and Robins [2020].

The distinction between forward causal inference and reverse attribution is important in interpreting this literature. The potential-outcomes framework is fundamentally organized around responses to specified interventions: the causal question concerns how an outcome would change under alternative treatments or exposures. Structural causal models likewise encode causal direction through structural equations and interventions. These frameworks therefore provide a natural language for asking what an intervention does to an outcome, but they do not imply that an observed outcome uniquely determines the intervention, treatment, or latent state that generated it. In general, multiple causal states may be compatible with the same realized outcome.

This is not a limitation of causal inference in the usual sense, since identification concerns whether a prespecified causal object can be recovered from the available experimental, observational, or structural information. Rather, it highlights a different identification problem. A reverse question begins with an observed consequence and asks which causal state or antecedent could have produced it. Without additional structural restrictions, such a question may admit multiple compatible explanations even when the forward causal effect is well defined and identified. The inverse problem is therefore not obtained simply by exchanging the labels of cause and effect: it requires an analysis of the information retained by the forward mechanism and of the additional information needed to make the reverse mapping unique.

Our question is related to identification but is logically different. Standard identification asks whether a specified causal functional is determined by the available distribution or experimental information. We instead ask a more structural question: *what information is intrinsically required to reverse a causal mechanism?* Given

$$Y = f(X, U),$$

the object of interest is not initially an average treatment effect or another functional of the data-generating distribution. It is the information about the latent state U that remains relevant for recovering X from Y.

This distinction matters because the full latent disturbance need not be the relevant inverse object. Two latent states may be observationally distinct as values of U while inducing exactly the same response to every intervention on X. Our response-function map

$$\Phi_f(u) = \big[x \mapsto f(x, u)\big]$$

therefore induces the equivalence relation

$$u \sim_f u' \quad \Longleftrightarrow \quad f(x, u) = f(x, u') \quad \text{for all } x.$$

The resulting quotient

$$Q = \Phi_f(U)$$

retains precisely the distinctions among latent states that matter for the structural response of Y to X.

The causal literature contains many notions of equivalence, invariance, and latent representation, but our use of the response-function quotient serves a different purpose. We use it to define the *minimal causal latent object for inverse recovery.* This gives a canonical separation between two sources of inverse difficulty: noninvertibility of the structural mechanism itself and lack of access to the latent information that remains causally relevant. The resulting decomposition is therefore complementary to conventional identification theory rather than a replacement for it.

2.2 Inverse problems and latent-variable identification

Inverse problems traditionally study the recovery of unobserved causes, parameters, or states from observed consequences. A central distinction in that literature is between existence, uniqueness, and stability of an inverse. Ill-posedness arises when the forward operator is noninjective, when its inverse is discontinuous, or when the available observations do not contain sufficient information for recovery. These ideas have also entered econometrics through nonparametric instrumental variables, measurement-error models, deconvolution, and other problems in which structural objects must be recovered from indirect observations; see, for example, Newey and Powell [2003], Hu [2008], Carrasco et al. [2007].

Our framework shares the inverse-problem perspective but identifies a specifically causal source of the inverse information requirement. The usual question is whether a forward operator can be inverted. Here the forward operator is itself a causal mechanism, and its latent state has an internal equivalence structure. Consequently, the relevant inverse problem is not necessarily

$$(Y, U) \longmapsto X.$$

Rather, after quotienting out latent distinctions that have no effect on the causal response, it is

$$(Y, Q) \longmapsto X.$$

This yields the decomposition

$$\text{inverse identification} = \text{structural invertibility} + \text{availability of the causal quotient}.$$

When the response $x \mapsto f(x, u)$ is injective but the quotient is unavailable, the inverse is structurally well defined but informationally inaccessible. Conversely, when the quotient is observed but the structural response is noninjective, no amount of additional information about the latent state can restore exact inversion. This separation between structural ambiguity and latent-information deficiency is central to our theory.

The distinction is particularly important in nonlinear latent-variable models. For example, if

$$Y = X + U^2,$$

the sign of U is irrelevant to the entire response function and therefore should not be counted as inverse information. The canonical quotient is $Q = U^2$. By contrast, in

$$Y = h(X) + U,$$

with h injective, the response-function map is injective in U, so the quotient retains the full disturbance. The two models can therefore have qualitatively different inverse information requirements even when both contain a one-dimensional latent disturbance.

This perspective is related to, but distinct from, the literature on causal inference with latent variables. That literature primarily studies how unobserved confounding, mediation, or other latent structure affects identification of causal effects or causal graphs. Our objective is instead to characterize the latent information that remains relevant for *reversing* a known causal mechanism. The distinction between latent causal structure and latent inverse information is fundamental to our approach.

2.3 Information theory and causal information

Information theory provides a natural language for quantifying uncertainty about latent variables and for studying the information revealed by observations. Mutual information, conditional mutual information, directed information, and related quantities have been used extensively to characterize causal dependence and causal influence, including approaches that translate causal deduction and causal effects into information-theoretic language; see, for example, Ay and Polani [2008], Janzing et al. [2013], Schamberg and Coleman [2020].

Our object is different. We do not measure the strength of a causal relationship between X and Y by mutual information. Instead, we first construct the causal quotient Q and then ask how much of the quotient remains unresolved after observing Y. This leads naturally to the conditional law

$$P(Q \in \cdot \mid Y)$$

and, in discrete settings, to quantities such as

$$H(Q \mid Y).$$

The conceptual distinction is important: information-theoretic dependence between X and Y is not the same object as information required to reverse the structural relationship between them.

For continuous variables, this distinction becomes especially consequential. Exact structural inversion can make conventional conditional mutual information unsuitable as a universal scalar measure of inverse difficulty. For example, if $Y = X + U$, then X is exactly determined by (Y, U), even though the relevant conditional-information quantities need not provide a finite measure of the amount of latent information that must be supplied. We therefore formulate the continuous theory in terms of conditional laws and information dimension rather than differential entropy alone.

The resulting quantity

$$\mathcal{R}_{\mathrm{IC}} = d_I(Q) - d_I(Q \mid Y)$$

may be interpreted as the number of intrinsic quotient dimensions revealed by the observation. Under the regularity conditions developed below, this information-theoretic object has a differential-geometric interpretation:

$$\mathcal{R}_{\mathrm{IC}} = r_{\mathrm{rev}}.$$

Thus the contribution is not a new information measure in the abstract. It is a causal interpretation and structural identification of a dimension drop that is otherwise purely information-theoretic.

2.4 Information dimension and rate-distortion theory

The use of information dimension connects our analysis to a separate information-theoretic literature. R'enyi's information dimension provides a notion of the intrinsic dimensionality of a random variable that remains meaningful for distributions that are not absolutely continuous with respect to ambient Lebesgue measure. Subsequent work has established deep connections between information dimension, high-resolution quantization, and rate-distortion theory; see Rényi [1959], Wu and Verdu [2010], Kawabata and Dembo [1994].

We use these results for a different purpose. The relevant random object is not the original latent disturbance U, whose dimension may depend on an arbitrary representation of the structural model, but the causal quotient Q. This makes the resulting dimension representation-invariant with respect to latent distinctions that do not affect the intervention-response function.

The distinction becomes particularly useful after conditioning on Y. The conditional information dimension

$$d_I(Q \mid Y = y)$$

describes the intrinsic dimension of the set of quotient states that remain compatible with the observation $Y = y$. Our geometric results show, under constant-rank and conditional-regularity assumptions, that this dimension equals

$$k - r_{\mathrm{rev}}.$$

Hence the information dimension of the unresolved inverse problem is determined by the number of quotient directions that remain after the observation has revealed r_{rev} independent directions.

This also yields an operational interpretation. We define an inverse-causal rate-distortion function by asking how much additional information about the quotient must be communicated, conditional on Y, to recover X to a prescribed distortion level. Under regularity,

$$R_{\mathrm{IC}}(\varepsilon) = \frac{d_I(Q \mid Y)}{2} \log \frac{1}{\varepsilon} + o\left(\log \frac{1}{\varepsilon}\right).$$

The high-resolution coefficient is therefore the conditional information dimension of the unresolved causal quotient. Existing rate-distortion theory supplies the information-theoretic machinery; our contribution is to identify the object being compressed as the causal quotient and to connect its rate-distortion dimension to the geometry of causal revelation.

2.5 Differential geometry, fibers, and conditional distributions

The geometric part of our argument uses standard tools from differential geometry, including the constant-rank theorem, regular level-set results, and the coarea formula. These results characterize the local structure and measure-theoretic decomposition of fibers of smooth maps; see, for example, Lee [2012], Federer [1969].

The novelty lies not in these geometric tools but in the particular fiber whose dimension is relevant to inverse causal information. For

$$Y = F(X, Q),$$

the observation fiber is

$$\mathcal{F}_y = \{(x, q) : F(x, q) = y\}.$$

However, the full fiber lives in the joint (X, Q)-space and therefore does not directly describe the ambiguity about Q. The relevant object is its projection onto the quotient space,

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y).$$

At a point (x, q), variation in Q may be observationally indistinguishable from variation in X. Consequently, the relevant derivative is not D_qF itself but its quotient modulo the directions generated by D_xF:

$$\overline{D}_qF : T_q\mathcal{Q} \longrightarrow T_y\mathcal{Y} / \operatorname{Im} D_xF.$$

We define

$$r_{\mathrm{rev}} = \operatorname{rank}(\overline{D}_qF).$$

Under the regularity conditions stated below, the projected observation fiber satisfies

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}}.$$

This construction provides the geometric core of the theory. The rank r_{rev} measures the number of locally distinguishable quotient directions that survive after the observational variation attributable to X has been factored out. It therefore has a direct causal interpretation that the ordinary Jacobian rank D_qF does not possess.

The coarea formula then supplies the bridge from this geometry to probability. Under a smooth joint density for (X, Q), conditional distributions on regular fibers admit Hausdorff-measure representations. After projection onto $\mathcal{Q}$, and under the conditional absolute-continuity and nondegeneracy assumptions imposed below, the law of $Q \mid Y = y$ is supported on the d-dimensional projected fiber

$$\mathcal{A}_y, \qquad d = k - r_{\text{rev}}.$$

Information dimension is then the appropriate probabilistic invariant of this conditional distribution. Thus the geometric and information-theoretic statements are two descriptions of the same unresolved inverse information.

2.6 Relation to existing notions of causal information flow

There is also a literature seeking geometric or information-theoretic measures of causal information flow. Such work includes approaches based on directed information, transfer entropy, conditional entropy, and geometric or fractal dimensions of dynamical systems. These approaches are useful for quantifying how information about one variable is transmitted to another. In particular, Surasinghe and Bollt [2020] develop a geometric information-flow measure based on conditional correlation dimension and interpret effective dimensionality as a geometric description of information flow.

Our objective differs in three respects. First, the relevant latent object is defined structurally through the response-function quotient rather than chosen directly from the observed variables. Second, revelation is defined relative to an inverse problem: a quotient direction is revealed only insofar as it reduces the ambiguity about the latent information needed to reconstruct the causal antecedent. Third, the resulting revelation rank is defined by a quotient derivative,

$$D_qF \mod \operatorname{Im} D_xF,$$

which explicitly removes observational variation that can be generated by changes in the variable being recovered.

Accordingly, our notion of causal revelation should not be interpreted as a replacement for mutual information, directed information, transfer entropy, or other measures of causal dependence. It answers a different question:

How many causally relevant latent degrees of freedom does the observation reveal for purposes of inverse recovery?

2.7 Contribution relative to the literature

The preceding literatures supply the components from which our theory is constructed, but none of them by itself provides the central object of the paper. The contribution can be summarized as the following chain:

$$U \xrightarrow{\Phi_f} Q \xrightarrow{F} Q \mid Y \longrightarrow \begin{cases} \text{projected-fiber dimension,} \\ \text{conditional information dimension,} \\ \text{inverse-causal rate-distortion.} \end{cases}$$

The response-function map determines which latent distinctions are causally relevant. The inverse-identification theorem separates structural noninvertibility from missing latent information.

The projected-fiber theorem identifies the local dimension of the unresolved quotient. Conditional information-dimension results convert that geometric dimension into an information-theoretic quantity. Finally, rate-distortion theory gives the resulting inverse problem an operational interpretation.

The principal identity is therefore

$$\boxed{d_I(Q \mid Y) = d_I(Q) - r_{\mathrm{rev}},}$$

under the regularity conditions developed in the paper. The left-hand side measures inverse difficulty, the first term on the right-hand side measures causal latent complexity, and the second measures causal revelation by the observation.

The paper's contribution is consequently best viewed not as a new estimator, a new causal-effect functional, or a new information measure in isolation, but as a theory of the information loss and information requirements generated by causal mechanisms. Existing causal identification, inverse-problem, information-theoretic, and differential-geometric results become components of this theory, while the causal quotient and the equality between geometric revelation and unresolved information dimension provide its central new objects and results.

3 Structural Setup and the Causal Quotient

We begin with a structural causal mechanism

$$Y = f(X, U), \tag{1}$$

where X is the causal variable whose value we seek to recover, Y is the observed outcome, and U collects latent disturbances. The central object in the analysis is not the particular representation of U, but the way in which each latent state changes the response of Y to interventions on X.

Throughout this section, let

$$(\mathcal{X}, X), \qquad (\mathcal{U}, U), \qquad (\mathcal{Y}, Y)$$

be measurable spaces, and let

$$f : \mathcal{X} \times \mathcal{U} \longrightarrow \mathcal{Y}$$

be measurable. We write X and U for random elements taking values in $\mathcal{X}$ and $\mathcal{U}$, respectively, and assume

$$Y = f(X, U) \qquad \text{a.s.}$$

3.1 Response functions

For each $u \in \mathcal{U}$, define the structural response function

$$f_u : \mathcal{X} \longrightarrow \mathcal{Y}, \qquad f_u(x) = f(x, u). \tag{2}$$

Thus $f_u(x)$ describes the outcome that would result from setting X to x while holding the latent state fixed at u.

The collection of all such response functions is generated by the map

$$\Phi_f : \mathcal{U} \longrightarrow \mathcal{Y}^{\mathcal{X}}, \qquad \Phi_f(u) = f_u. \tag{3}$$

The codomain $\mathcal{Y}^{\mathcal{X}}$ is equipped with the evaluation sigma-field

$$E = \sigma\left\{ e_x^{-1}(B) : x \in \mathcal{X},\ B \in Y \right\},$$

where

$$e_x : \mathcal{Y}^{\mathcal{X}} \longrightarrow \mathcal{Y}, \qquad e_x(h) = h(x).$$

Because

$$e_x \circ \Phi_f(u) = f(x, u),$$

measurability of f implies measurability of Φ_f.

The map Φ_f is the fundamental structural object of the paper. It records the entire intervention-response function generated by a latent state rather than only the realized outcome under one value of X.

3.2 Structural causal equivalence

Two latent states need not be distinct for purposes of the causal mechanism. We therefore define the relation

$$u \sim_f u' \iff f(x, u) = f(x, u') \quad \text{for every } x \in \mathcal{X}. \tag{4}$$

Thus $u \sim_f u'$ if and only if the two latent states induce exactly the same structural response to every intervention on X.

[Causal equivalence] The relation $\sim_f$ defined in (4) is called *structural causal equivalence*. Its equivalence class is denoted

$$[u]_f = \{u' \in \mathcal{U} : u' \sim_f u\}.$$

The quotient space

$$\mathcal{U}/{\sim_f}$$

is called the *causal quotient* of the latent state space induced by f.

The terminology emphasizes that equivalence is defined by the full intervention-response function. In particular, it is stronger than equality of the realized outcome $f(X, U)$, equality of conditional distributions of Y under a particular observational distribution, or any other equivalence defined from a single observational regime.

The response-function map provides a concrete representation of this quotient. Indeed,

$$\Phi_f(u) = \Phi_f(u') \iff u \sim_f u'. \tag{5}$$

Consequently, the fibers of Φ_f are exactly the causal equivalence classes. The quotient can therefore be represented by the image

$$\Phi_f(\mathcal{U}) \subseteq \mathcal{Y}^{\mathcal{X}}.$$

For the random latent state U, define the *causal quotient random element*

$$Q =$$

$\Phi_f(U)$.(6) The random variable Q records the structural response function associated with the realized latent state.

[Canonical representation of the causal quotient] Let $Q = \Phi_f(U)$. Then the following statements hold.

1. Q is measurable with respect to U: $\sigma(Q) \subseteq \sigma(U)$.

“‘

2. Two latent states have the same value of Q if and only if they are structurally causally equivalent:
$$Q(u) = Q(u') \iff u \sim_f u'.$$
3. Any measurable representation $S = s(U)$ that preserves the complete structural response function must factor through Q, subject to the usual measurable factorization conditions. In particular, if
$$\Phi_f(U) = \psi(S) \qquad \text{a.s.}$$
for some measurable ψ, then
$$Q = \psi(S) \qquad \text{a.s.}$$
"‘

Proof. The first statement follows directly from the measurability of Φ_f:

$$Q = \Phi_f(U).$$

For the second statement,

$$Q(u) = Q(u')$$

if and only if

$$\Phi_f(u) = \Phi_f(u'),$$

which, by definition of Φ_f, is equivalent to

$$f(x,u) = f(x,u') \qquad \text{for every } x \in \mathcal{X}.$$

This is precisely $u \sim_f u'$.

The third statement follows immediately from the assumed factorization

$$\Phi_f(U) = \psi(S).$$

Since $Q = \Phi_f(U)$, we have

$$Q = \psi(S) \qquad \text{a.s.}$$

The nontrivial issue is therefore not the existence of the pointwise factorization but the measurable realization of the quotient. Under standard-Borel assumptions, the required measurable factorization is available whenever S is constant on the fibers of Φ_f. □

3.3 The causal quotient as the relevant latent object

The quotient construction separates latent complexity from causal complexity. The dimension or entropy of U can depend on how the structural model chooses to parameterize latent states. The quotient Q, by contrast, removes distinctions that have no effect on the response of Y to interventions on X.

Consider first the additive model
$$Y = h(X) + U, \tag{7}$$
where h is known. The response function is

$$f_u(x) = h(x) + u.$$

If $u \neq u'$, then

$$f_u(x) \neq f_{u'}(x)$$

for every x, and hence

$$u \sim_f u' \quad \Longleftrightarrow \quad u = u'.$$

The response-function map is therefore injective, so Q contains the same information as U.

Now consider

$$Y = X + U^2. \tag{8}$$

Here

$$f_u(x) = x + u^2.$$

Consequently,

$$f_u = f_{u'} \quad \Longleftrightarrow \quad u^2 = (u')^2,$$

and therefore

$$u \sim_f u' \quad \Longleftrightarrow \quad u' = \pm u.$$

The causal quotient is represented by

$$Q = U^2.$$

The sign of U may be part of the latent representation, but it has no effect on any intervention response. It is therefore not part of the causal information required to reverse the mechanism.

More generally, consider

$$Y = X + U_1 + U_2^2. \tag{9}$$

The response function is

$$f_{(u_1,u_2)}(x) = x + u_1 + u_2^2.$$

Hence two latent states are equivalent whenever

$$u_1 + u_2^2 = u_1' + (u_2')^2.$$

The canonical quotient is therefore

$$Q = U_1 + U_2^2. \tag{10}$$

This example illustrates why the quotient should be defined through Φ_f, rather than by selecting a collection of apparently relevant latent coordinates in advance.

3.4 Causal information loss

The map

$$U \xrightarrow{\Phi_f} Q \tag{11}$$

defines a structural compression of the latent state. Information about U that is discarded by this map corresponds exactly to distinctions among latent states that do not affect the structural response to X.

At the level of sigma-fields,

$$\boxed{\sigma(Q) \subseteq \sigma(U).} \tag{12}$$

We refer to the passage from U to Q as *causal information loss*.

For discrete random variables, this loss admits the direct information-theoretic representation

$$\mathrm{CIL}(U; f) =======================$$

H(U| Q)
H(U)-H(Q), (13)

where the second equality follows because Q is a deterministic function of U.

The quantity in (3.4) should not be interpreted as the information required to perform inverse recovery. It measures a different phenomenon: information about the latent representation that is *causally redundant*. The inverse problem concerns the information in Q that remains unavailable after observing Y.

This distinction gives rise to two conceptually separate transformations:

$$U \longrightarrow Q \longrightarrow Q \mid Y. \tag{14}$$

The first transformation removes latent distinctions that are causally irrelevant. The second describes the causal information that remains unresolved after observing the outcome.

The first is *causal information loss*; the second is *inverse information deficiency*. The remainder of the paper studies the second object while using the first to define the appropriate latent state space.

3.5 Structural inversion after quotienting

The quotient construction also clarifies what it means for a causal mechanism to be invertible.

Suppose that for P_U-almost every u, the map

$$x \longmapsto f_u(x)$$

is injective. Then, whenever the relevant response function is known, the structural inverse is well defined on its image:

$$x = f_u^{-1}(y). \tag{15}$$

Since f_u is constant on causal equivalence classes, the inverse can be represented in terms of the quotient:

$$X = g(Y, Q) \qquad \text{a.s.} \tag{16}$$

for a suitable measurable inverse map g, under the regularity conditions specified below.

Equation (16) makes the role of the quotient precise. The original latent variable U is not itself necessarily needed for inversion. What is needed is the equivalence class of U under equality of structural response functions.

This observation leads to the central decomposition of inverse identification developed in the next section. There are two logically distinct ways in which inversion can fail. First, the response function f_u may not be injective in x, in which case distinct values of X produce the same outcome even when the entire latent state is known. Second, the response function may be invertible, but the relevant quotient Q may not be available from the information observed by the researcher.

The remainder of the paper treats these two sources of ambiguity separately.

4 Projected-Fiber Geometry and Causal Revelation

The previous section established that, once the causal quotient Q is known, inverse recovery is a structural question. We now characterize how much of Q remains unresolved after observing Y.

The appropriate object is the observation fiber in the joint (X, Q)-space and, crucially, its projection onto the quotient space. The distinction between these two objects is what allows the geometry of inverse ambiguity to be separated from variation in the causal variable itself.

Throughout this section, let

$$\mathcal{X}, \quad \mathcal{Q}, \quad \mathcal{Y}$$

be C^r manifolds, with $r \geq 1$, of dimensions

$$\dim \mathcal{X} = p, \qquad \dim \mathcal{Q} = k, \qquad \dim \mathcal{Y} = m.$$

Let

$$F : \mathcal{X} \times \mathcal{Q} \longrightarrow \mathcal{Y}$$

be a C^r map, and consider the structural mechanism

$$Y = F(X, Q). \tag{17}$$

4.1 Observation fibers and projected fibers

For a given $y \in \mathcal{Y}$, define the observation fiber

$$\mathcal{F}_y = F^{-1}(y) = \{(x, q) \in \mathcal{X} \times \mathcal{Q} : F(x, q) = y\}. \tag{18}$$

The fiber $\mathcal{F}_y$ contains all pairs (x, q) that could have generated the observed outcome y. It therefore describes the full observational ambiguity in the joint state.

For inverse recovery of X, however, the entire fiber is not the relevant object. What matters is which quotient states q remain compatible with y, allowing x to vary freely within its structural support. Let

$$\pi_Q : \mathcal{X} \times \mathcal{Q} \longrightarrow \mathcal{Q}$$

denote the canonical projection,

$$\pi_Q(x, q) = q.$$

We define the *projected observation fiber*

$$\boxed{\mathcal{A}_y = \pi_Q(\mathcal{F}_y) = \{q \in \mathcal{Q} : \exists x \in \mathcal{X},\ F(x, q) = y\}.} \tag{19}$$

Thus $\mathcal{A}_y$ is the set of causal quotient states that remain compatible with the observation $Y = y$.

This distinction is fundamental. The dimension of $\mathcal{F}_y$ measures ambiguity jointly in X and Q. The dimension of $\mathcal{A}_y$ measures ambiguity specifically in the causal information relevant to inverse recovery.

4.2 The quotient derivative

The next step is to determine which directions in Q are revealed by a local change in the outcome.

At a point $(x,q) \in \mathcal{F}_y$, consider the differential

$$DF_{(x,q)} : T_x\mathcal{X} \oplus T_q\mathcal{Q} \longrightarrow T_y\mathcal{Y}.$$

Write

$$D_xF : T_x\mathcal{X} \longrightarrow T_y\mathcal{Y}$$

and

$$D_qF : T_q\mathcal{Q} \longrightarrow T_y\mathcal{Y}$$

for the corresponding partial derivatives.

A change in q is not necessarily observable as a change in y, because the same change in y may be generated by changing x. The directions generated by X are therefore nuisance directions for the purpose of identifying Q.

Define

$$\mathcal{S}_X(x,q) = \operatorname{Im} D_xF(x,q) \subseteq T_y\mathcal{Y}. \tag{20}$$

The relevant derivative of F with respect to Q is the induced map

$$\boxed{\overline{D}_qF(x,q) : T_q\mathcal{Q} \longrightarrow T_y\mathcal{Y}/\mathcal{S}_X(x,q),} \tag{21}$$

defined by

$$\overline{D}_qF(x,q)(v_q) = D_qF(x,q)v_q + \mathcal{S}_X(x,q). \tag{22}$$

In words, $\overline{D}_qF$ records the component of a quotient perturbation that cannot be replicated, to first order, by a perturbation of X.

[Causal revelation rank] At a point (x,q), the *causal revelation rank* is

$$\boxed{r_{\text{rev}}(x,q) = \operatorname{rank} \overline{D}_qF(x,q).} \tag{23}$$

The terminology reflects the inverse interpretation. A quotient direction belongs to the kernel of $\overline{D}_qF$ precisely when its first-order effect on Y can be offset by some first-order change in X. Such a direction is therefore locally unresolved by the observation.

More explicitly,

$$\ker \overline{D}_qF(x,q) = \{v_q \in T_q\mathcal{Q} : \exists v_x \in T_x\mathcal{X} \text{ such that } D_xF(x,q)v_x + D_qF(x,q)v_q = 0\}. \tag{24}$$

Thus the unresolved quotient tangent space is

$$\mathcal{N}_q(x,q) = \ker \overline{D}_qF(x,q).$$

4.3 Projected-Fiber Geometry Theorem

We now give the principal geometric result.

[Projected-Fiber Geometry] Let

$$F : \mathcal{X} \times \mathcal{Q} \to \mathcal{Y}$$

be a C^r map, $r \geq 1$, with

$$\dim \mathcal{X} = p, \qquad \dim \mathcal{Q} = k.$$

Let $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Q}$ be a F-invariant structural support, and fix $y \in \mathcal{Y}$. Suppose that on a neighborhood of

$$\mathcal{F}_y \cap \mathcal{S}$$

the following conditions hold:

(A1) F has constant rank r_F;

(A2) $D_x F$ has constant rank r_X;

(A3) the induced map

$$\overline{D}_q F : T_q \mathcal{Q} \to T_y \mathcal{Y} / \operatorname{Im} D_x F$$

has constant rank

$$r_{\text{rev}};$$

(A4) the restriction

$$\pi_Q \big|_{\mathcal{F}_y \cap \mathcal{S}}$$

is a proper embedding onto its image.

Then $\mathcal{F}_y \cap \mathcal{S}$ is a properly embedded submanifold of $\mathcal{X} \times \mathcal{Q}$, and

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y \cap \mathcal{S})$$

is a properly embedded submanifold of $\mathcal{Q}$ satisfying

$$\boxed{\dim \mathcal{A}_y = k - r_{\text{rev}}.} \tag{25}$$

Equivalently, the local dimension of the unresolved causal quotient equals the dimension of the quotient space minus the number of quotient directions revealed by the observation.

Proof. We proceed in four steps.

Step 1: geometry of the observation fiber. Since F has constant rank r_F on a neighborhood of the relevant support, the constant-rank level-set theorem implies that

$$\mathcal{F}_y \cap \mathcal{S} = F^{-1}(y) \cap \mathcal{S}$$

is a submanifold of $\mathcal{X} \times \mathcal{Q}$ of dimension

$$\dim(\mathcal{F}_y \cap \mathcal{S}) = p + k - r_F. \tag{26}$$

At every $(x, q) \in \mathcal{F}_y \cap \mathcal{S}$,

$$T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S}) = \ker DF_{(x,q)}.$$

Step 2: differential of the quotient projection. Consider the restriction

$$\pi_Q\big|_{\mathcal{F}_y \cap \mathcal{S}}.$$

Its differential at (x, q) is simply the restriction of the canonical projection:

$$D\pi_Q(v_x, v_q) = v_q.$$

Therefore,

$$\begin{aligned} D\pi_Q\left(T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S})\right) &= \{v_q : \exists v_x, (v_x, v_q) \in \ker DF\} && (27) \\ &= \{v_q : \exists v_x, D_x F v_x + D_q F v_q = 0\}. && (28) \end{aligned}$$

By definition of the quotient derivative,

$$D_x F v_x + D_q F v_q = 0$$

for some v_x if and only if

$$D_q F v_q \in \operatorname{Im} D_x F,$$

which is equivalent to

$$\overline{D}_q F(v_q) = 0.$$

Consequently,

$$D\pi_Q\left(T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S})\right) = \ker \overline{D}_q F. \tag{29}$$

Step 3: dimension of the projected tangent space. By the rank-nullity theorem applied to

$$\overline{D}_q F : T_q\mathcal{Q} \to T_y\mathcal{Y}/\operatorname{Im} D_x F,$$

we have

$$\dim \ker \overline{D}_q F = k - r_{\text{rev}}. \tag{30}$$

Combining (29) and (30),

$$\operatorname{rank} D\left(\pi_Q\big|_{\mathcal{F}_y \cap \mathcal{S}}\right) = k - r_{\text{rev}}. \tag{31}$$

Assumption (A4) states that this restricted projection is an embedding onto its image. In particular, its image is an embedded submanifold and its dimension equals the rank of its differential. Hence

$$\dim \mathcal{A}_y = k - r_{\text{rev}}.$$

This proves (25). □

4.4 An equivalent rank formula

The revelation rank admits an alternative expression that makes the geometry particularly transparent.

Since

$$DF_{(x,q)} = \begin{bmatrix} D_x F & D_q F \end{bmatrix},$$

the image of DF is generated by the images of $D_x F$ and $D_q F$. Therefore

$$\operatorname{rank} DF = r_X + r_{\text{rev}}. \tag{32}$$

Hence

$$\boxed{r_{\text{rev}} = \operatorname{rank} DF - \operatorname{rank} D_x F.} \tag{33}$$

This identity gives an especially simple interpretation of causal revelation. The full outcome map has rank

$$\operatorname{rank} DF,$$

but some of this rank is already generated by variation in X. The remainder is the number of locally distinguishable directions supplied by Q beyond those generated by X.

Substituting (32) into the dimension formula yields

$$\dim \mathcal{A}_y = k - \operatorname{rank} DF + \operatorname{rank} D_x F. \tag{34}$$

The same expression can be obtained directly from the tangent-space dimension. Indeed,

$$\dim \ker DF = p + k - \operatorname{rank} DF.$$

The kernel of the restricted projection

$$D\pi_Q : \ker DF \to T_q \mathcal{Q}$$

consists of vectors of the form $(v_x, 0)$ satisfying

$$D_x F v_x = 0.$$

Its dimension is therefore

$$p - r_X.$$

Consequently,

$$\begin{aligned} \operatorname{rank} D\left(\pi_Q\big|_{\mathcal{F}_y}\right) &= (p + k - r_F) - (p - r_X) && (35)\\ &= k - r_F + r_X && (36)\\ &= k - r_{\text{rev}}, && (37) \end{aligned}$$

as required.

4.5 Interpretation as causal revelation

The theorem gives a local geometric meaning to the phrase "information revealed by Y." Suppose that Q has dimension k. If

$$r_{\text{rev}} = 0,$$

then

$$\dim \mathcal{A}_y = k.$$

Locally, observing $Y = y$ does not reduce the dimension of the quotient states compatible with the outcome. Every quotient direction can be offset, to first order, by a suitable change in X.

At the other extreme, if

$$r_{\text{rev}} = k,$$

then

$$\dim \mathcal{A}_y = 0.$$

Locally, the observation identifies the quotient state up to zero-dimensional ambiguity. Under additional global conditions, this can lead to local or global identification of Q.

The intermediate case

$$0 < r_{\mathrm{rev}} < k$$

is the generic partial-revelation case. The observation eliminates r_{rev} quotient dimensions while leaving

$$k - r_{\mathrm{rev}}$$

dimensions unresolved.

This gives the following interpretation:

$$\boxed{\text{causal latent complexity} = \text{revealed quotient dimension} + \text{unresolved quotient dimension.}} \tag{38}$$

In the regular k-dimensional setting,

$$k = r_{\mathrm{rev}} + \dim \mathcal{A}_y.$$

4.6 Examples

No quotient revelation. Consider

$$Y = X + Q, \tag{39}$$

with

$$X, Q \in \mathbb{R}.$$

Then

$$D_x F = 1, \qquad D_q F = 1.$$

Because

$$\operatorname{Im} D_x F = T_y \mathbb{R},$$

the quotient derivative is zero:

$$\overline{D}_q F = 0.$$

Hence

$$r_{\mathrm{rev}} = 0, \qquad \dim \mathcal{A}_y = 1.$$

Indeed,

$$\mathcal{A}_y = \{q : y = x + q \text{ for some } x\} = \mathbb{R}.$$

Knowing $Y = y$ does not reduce the set of possible values of Q at all when X is unrestricted.

Complete quotient revelation. Consider

$$Y = (X, Q), \tag{40}$$

with

$$X \in \mathbb{R}^p, \qquad Q \in \mathbb{R}^k.$$

The image of D_xF occupies the X-coordinates of the outcome, while the quotient derivative spans the Q-coordinates. Thus

$$r_{\text{rev}} = k.$$

The projected fiber is zero-dimensional:

$$\dim \mathcal{A}_y = 0.$$

Indeed, Q is directly revealed by Y.

Partial revelation. Consider

$$Y = (X_1 + Q_1, Q_2), \tag{41}$$

where

$$X = (X_1, X_2) \in \mathbb{R}^2, \qquad Q = (Q_1, Q_2) \in \mathbb{R}^2.$$

The first quotient direction Q_1 is confounded with variation in X_1, while the second quotient direction Q_2 enters an outcome coordinate independently of X. Hence

$$r_{\text{rev}} = 1.$$

The theorem gives

$$\dim \mathcal{A}_y = 2 - 1 = 1.$$

Exactly one quotient dimension remains unresolved.

4.7 Local versus global revelation

The rank r_{rev} is intrinsically a local object. It describes the dimension of the tangent space of the projected fiber and therefore characterizes local distinguishability.

It does not, by itself, guarantee global identification of Q. A zero-dimensional projected fiber may contain several isolated points, for example when the structural map has discrete symmetries. Conversely, a positive-dimensional projected fiber can have complicated global topology even when its local dimension is constant.

The theorem therefore separates two questions:

1. *How many continuous quotient directions remain locally unresolved?*
2. *How many globally distinct quotient states remain compatible with the observation?*

The first is answered by r_{rev}. The second requires additional global restrictions, such as injectivity, connectedness, or support restrictions. This distinction will become important when we pass from geometric dimension to conditional information dimension.

4.8 Why the projection is essential

One might instead attempt to characterize inverse ambiguity using the dimension of the full fiber

$$\mathcal{F}_y.$$

That quantity is generally not the relevant object.

For example, if

$$Y = X + Q$$

with $X, Q \in \mathbb{R}$, then

$$\dim \mathcal{F}_y = 1.$$

But this one-dimensional fiber represents joint variation in (X, Q). Its dimension alone does not tell us whether the unresolved variation concerns X, Q, or both.

The projection

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y)$$

removes this ambiguity. It asks directly which quotient states remain possible after allowing X to vary.

The quotient derivative

$$\overline{D}_q F$$

then performs the corresponding infinitesimal operation: it removes outcome directions already generated by X and measures only the independent outcome variation attributable to Q.

Thus the pair

$$\left(\mathcal{A}_y, \overline{D}_q F\right)$$

provides the geometric representation of inverse causal ambiguity.

4.9 From geometry to probability

The theorem is deterministic. It characterizes the dimension of the set of quotient states compatible with a particular outcome. To obtain an information-theoretic interpretation, we must connect this geometric set to the conditional probability law

$$P(Q \in \cdot \mid Y = y).$$

Geometry alone is insufficient for this step. A probability measure may be concentrated on a lower-dimensional subset of a geometrically larger projected fiber. We therefore require a conditional nondegeneracy condition ensuring that the conditional law occupies the full-dimensional portion of the projected fiber.

In the next section, we impose the corresponding regularity conditions and show that, under conditional absolute continuity with respect to Hausdorff measure,

$$\boxed{d_I(Q \mid Y = y) = \dim \mathcal{A}_y = k - r_{\mathrm{rev}}}$$

for almost every y.

The projected-fiber theorem therefore provides the geometric half of the paper's central identity. The next section supplies the probabilistic half.

5 Conditional Information-Dimension Revelation

The projected-fiber geometry of the previous section is deterministic. It characterizes, locally, the set of quotient states that remain compatible with an observed outcome. To turn this geometric statement into an information-theoretic one, we now place a probability law on the structural variables and study the conditional distribution of the causal quotient given the observed outcome.

The key point is that, under appropriate regularity, the dimension of the projected fiber is also the information dimension of the conditional quotient distribution. Thus the rank of the quotient derivative becomes a quantitative measure of causal information revealed by the outcome.

5.1 Conditional quotient distributions

Let (X, Q) be a random element of $\mathcal{X} \times \mathcal{Q}$ with probability law μ, and let

$$Y = F(X, Q)$$

with law

$$\nu = F_{\#}\mu.$$

Assume that the relevant measurable spaces are standard Borel so that regular conditional distributions exist. Write

$$\mu_y = P_{X,Q|Y=y}$$

for a regular conditional distribution of (X, Q) given $Y = y$, defined for ν-almost every y.
Let

$$\pi_Q : \mathcal{X} \times \mathcal{Q} \to \mathcal{Q}$$

denote projection onto the quotient coordinate. The conditional law of the quotient given the outcome is then

$$\lambda_y = (\pi_Q)_{\#}\mu_y = P_{Q|Y=y}.$$

The projected fiber introduced above is

$$\mathcal{A}_y = \pi_Q\big(F^{-1}(y) \cap \mathcal{S}\big),$$

where $\mathcal{S}$ denotes the relevant structural support. Since μ_y is concentrated on the level set $F^{-1}(y)$ for ν-almost every y, it follows immediately that

$$\boxed{\operatorname{supp}(\lambda_y) \subseteq \mathcal{A}_y} \qquad \nu\text{-a.s.}$$

Thus the projected fiber is not merely a deterministic compatibility set: it contains the support of the posterior distribution over causally relevant latent states.

Suppose, in the constant-rank regime of Section 4, that

$$r_{\mathrm{rev}} = \operatorname{rank} \overline{D}_q F$$

is constant on the relevant structural support. Define

$$d = k - r_{\mathrm{rev}}.$$

The geometric theorem then gives

$$\dim \mathcal{A}_y = d.$$

To connect this dimension to information dimension, we impose a conditional regularity condition.

[Conditional geometric regularity] For ν-almost every y, the projected fiber $\mathcal{A}_y$ is a d-dimensional embedded submanifold and

$$\lambda_y \ll \mathcal{H}^d \lceil_{\mathcal{A}_y} .$$

Moreover, writing

$$\rho_y = \frac{d\lambda_y}{d(\mathcal{H}^d \lceil_{\mathcal{A}_y})},$$

the density is positive and finite $\mathcal{H}^d$-almost everywhere and satisfies the usual local regularity conditions required for information dimension to be invariant under local bi-Lipschitz parametrization.

The assumption separates two logically distinct facts. The geometric theorem determines the dimension of the compatibility set $\mathcal{A}_y$. The probabilistic assumption ensures that the conditional distribution actually occupies that set in a nondegenerate d-dimensional way. Without the latter condition, the support may be d-dimensional while the probability law is concentrated on a lower-dimensional subset.

5.2 Conditional information dimension

For a random vector $Q \in \mathbb{R}^k$, define the quantized quotient

$$[Q]_\delta = \delta \left\lfloor \frac{Q}{\delta} \right\rfloor$$

componentwise. The conditional lower and upper information dimensions are

$$\underline{d}_I(Q \mid Y) = \liminf_{\delta \downarrow 0} \frac{H([Q]_\delta \mid Y)}{\log(1/\delta)},$$

and

$$\overline{d}_I(Q \mid Y) = \limsup_{\delta \downarrow 0} \frac{H([Q]_\delta \mid Y)}{\log(1/\delta)},$$

whenever the conditional entropies are well defined. When the two limits coincide, their common value is denoted

$$d_I(Q \mid Y).$$

For a fixed y, the analogous quantity is the information dimension of the conditional measure λ_y:

$$d_I(\lambda_y) = \lim_{\delta \downarrow 0} \frac{H_{\lambda_y}([Q]_\delta)}{\log(1/\delta)},$$

provided the limit exists.

The following theorem is the central bridge between the differential geometry of the causal mechanism and the information dimension of inverse uncertainty.

[Conditional Information-Dimension Revelation] Let

$$Y = F(X, Q),$$

where $\mathcal{X}$, $\mathcal{Q}$, and $\mathcal{Y}$ are C^r manifolds with

$$\dim \mathcal{X} = p, \qquad \dim \mathcal{Q} = k.$$

Suppose that, on the relevant structural support, the hypotheses of the projected-fiber theorem hold with constant quotient-revelation rank

$$r_{\mathrm{rev}} = \operatorname{rank} \overline{D}_q F.$$

Let

$$d = k - r_{\mathrm{rev}}.$$

Suppose further that for ν-almost every y the conditional quotient law

$$\lambda_y = P_{Q|Y=y}$$

is absolutely continuous with respect to

$$\mathcal{H}^d \restriction_{\mathcal{A}_y},$$

with a nondegenerate density satisfying the regularity conditions of Assumption 5.1.
Then, for ν-almost every y,

$$\boxed{d_I(Q \mid Y = y) = d = k - r_{\mathrm{rev}}.}$$

Consequently, whenever the unconditional quotient has information dimension

$$d_I(Q) = k,$$

the amount of quotient information revealed by observing Y is

$$\boxed{d_I(Q) - d_I(Q \mid Y) = r_{\mathrm{rev}}.}$$

Proof. By the projected-fiber theorem,

$$\mathcal{A}_y = \pi_Q(F^{-1}(y) \cap \mathcal{S})$$

is, for ν-almost every y, a d-dimensional embedded submanifold with

$$d = k - r_{\mathrm{rev}}.$$

The conditional quotient law satisfies

$$\lambda_y(\mathcal{A}_y) = 1.$$

By assumption,

$$\lambda_y \ll \mathcal{H}^d \restriction_{\mathcal{A}_y},$$

with a nondegenerate density. Hence, locally on $\mathcal{A}_y$, the conditional law can be represented in a d-dimensional coordinate chart.

Let

$$\psi_y : V_y \subseteq \mathbb{R}^d \to \mathcal{A}_y$$

be a local C^1 parametrization. After restricting to sufficiently small coordinate neighborhoods, ψ_y and its inverse are bi-Lipschitz. The pullback measure

$$\widetilde{\lambda}_y = (\psi_y^{-1})_\# \lambda_y$$

is therefore an absolutely continuous d-dimensional probability measure with a nondegenerate density on the corresponding coordinate domain.

For a regular absolutely continuous measure on $\mathbb{R}^d$, information dimension is d. Moreover, information dimension is invariant under bi-Lipschitz transformations. Therefore the restriction of λ_y to each regular coordinate neighborhood has information dimension d.

An exhaustion of $\mathcal{A}_y$ by such coordinate neighborhoods, together with the assumed regularity of the conditional density, yields

$$d_I(\lambda_y) = d.$$

Since $\lambda_y = P_{Q|Y=y}$,

$$d_I(Q \mid Y = y) = d = k - r_{\text{rev}}.$$

If the unconditional quotient satisfies

$$d_I(Q) = k,$$

then subtracting the conditional dimension gives

$$d_I(Q) - d_I(Q \mid Y) = k - (k - r_{\text{rev}}) = r_{\text{rev}}.$$

This proves the result. □

5.3 The revelation identity

The theorem yields a particularly simple interpretation of the quotient derivative. Define the *causal revelation dimension* by

$$\boxed{\mathcal{R}_{\text{IC}} = d_I(Q) - d_I(Q \mid Y),}$$

whenever the two information dimensions exist. Under the conditions of Theorem 5.2 and a full-dimensional unconditional quotient,

$$\boxed{\mathcal{R}_{\text{IC}} = r_{\text{rev}}.}$$

Thus the differential rank introduced in Section 4 is not merely a local geometric diagnostic. It is the infinitesimal counterpart of the number of quotient dimensions statistically eliminated by observation.

The corresponding decomposition is

$$\boxed{k = \underbrace{r_{\text{rev}}}_{\text{revealed quotient dimension}} + \underbrace{(k - r_{\text{rev}})}_{\text{unresolved quotient dimension}}.}$$

Equivalently,

$$\boxed{\text{causal latent complexity} = \text{causal revelation} + \text{inverse uncertainty.}}$$

This gives a precise sense in which inverse difficulty is generated by information that remains latent after the outcome has been observed.

5.4 Why the result is causal rather than merely observational

The quantity

$$d_I(Q \mid Y)$$

is observational: it describes the conditional distribution of the quotient after seeing Y. By itself, it does not define what counts as causally relevant latent information.

That role is played by the response-function map

$$\Phi_f : U \to \mathcal{F}(\mathcal{X}, \mathcal{Y})$$

and the induced quotient

$$Q = \Phi_f(U).$$

Only distinctions in U that alter the causal response function survive into Q. Consequently,

$$\Phi_f \quad \text{determines the causal state space,}$$

whereas

$$P(Q \mid Y) \quad \text{determines how much of that causal state space is revealed by observation.}$$

This distinction is essential. An observationally sufficient representation of Y need not preserve the intervention-response structure of the latent disturbance. The present construction first quotients latent states according to their causal consequences and only then asks how observation reduces uncertainty over those causally distinct states.

The resulting sequence is therefore

$$\boxed{U \longrightarrow Q = \Phi_f(U) \longrightarrow P(Q \mid Y) \longrightarrow d_I(Q \mid Y).}$$

The first arrow removes causally redundant latent distinctions. The second incorporates observational information. The third converts the remaining uncertainty into an intrinsic dimensional quantity.

5.5 Coarea representation

The conditional absolute-continuity assumption can be motivated directly from smooth probability models through the coarea formula.

Suppose, for simplicity, that

$$\mathcal{X} \times \mathcal{Q}$$

and $\mathcal{Y}$ are Euclidean spaces and that (X, Q) has a sufficiently regular joint density

$$p_{X,Q}(x, q).$$

Let

$$r_F = \operatorname{rank} DF$$

be constant on the relevant support. The coarea formula gives, schematically,

$$\int_{\mathcal{X}\times\mathcal{Q}} \varphi(x,q) J_{r_F} F(x,q)\, d(x,q) = \int_{\mathcal{Y}} \int_{F^{-1}(y)} \varphi(x,q)\, d\mathcal{H}^{p+k-r_F}(x,q)\, dy,$$

where $J_{r_F}F$ denotes the appropriate r_F-dimensional Jacobian.

Consequently, under the usual nondegeneracy conditions, the conditional law of (X, Q) given $Y = y$ has a density on the level set

$$F^{-1}(y)$$

proportional to

$$\frac{p_{X,Q}(x, q)}{J_{r_F}F(x, q)}$$

with respect to the Hausdorff measure on that level set.

The quotient conditional law is obtained by pushing this measure forward under

$$\pi_Q.$$

When the restricted projection

$$\pi_Q\big|_{F^{-1}(y)\cap\mathcal{S}}$$

is a proper embedding, as in the projected-fiber theorem, the projected measure inherits a d-dimensional geometric structure on

$$\mathcal{A}_y, \qquad d = k - r_{\mathrm{rev}}.$$

Under nondegenerate densities, this gives precisely the conditional absolute continuity required by Theorem 5.2.

The coarea formula is therefore a probabilistic bridge between the ambient joint density and the lower-dimensional conditional quotient distribution. It does not itself generate the causal result: the causal content enters through the quotient construction and the projected-fiber geometry.

5.6 Examples

Additive confounding. Consider

$$Y = X + Q, \qquad X, Q \in \mathbb{R}.$$

Then

$$D_xF = 1, \qquad D_qF = 1,$$

and therefore

$$\operatorname{Im} D_xF = \mathbb{R}.$$

The quotient derivative takes values in

$$T_y\mathcal{Y}/\operatorname{Im} D_xF = \{0\},$$

so

$$r_{\mathrm{rev}} = 0.$$

The projected fiber is

$$\mathcal{A}_y = \{q : y - q \in \mathcal{X}\},$$

which is locally one-dimensional. Under a regular continuous distribution,

$$d_I(Q \mid Y = y) = 1.$$

If $d_I(Q) = 1$, then

$$d_I(Q) - d_I(Q \mid Y) = 0.$$

Observation of Y reveals no quotient dimension: the change in Y induced by Q is locally indistinguishable from a change in X.

Direct observation of the quotient. Now consider

$$Y = (X, Q).$$

The quotient perturbation changes the second outcome coordinate in a direction that cannot be generated by changing X. Hence

$$r_{\text{rev}} = k.$$

The projected fiber is locally zero-dimensional:

$$\dim \mathcal{A}_y = 0.$$

Thus

$$d_I(Q \mid Y = y) = 0.$$

If $d_I(Q) = k$,

$$d_I(Q) - d_I(Q \mid Y) = k.$$

The observation reveals the entire quotient.

Partial revelation. Let

$$X = (X_1, X_2), \qquad Q = (Q_1, Q_2),$$

and

$$Y = (X_1 + Q_1, Q_2).$$

The first quotient direction Q_1 is confounded with X_1, while the second quotient direction Q_2 is directly observed. Therefore

$$r_{\text{rev}} = 1.$$

Since $k = 2$,

$$d_I(Q \mid Y = y) = 1.$$

For a full-dimensional quotient,

$$d_I(Q) - d_I(Q \mid Y) = 2 - 1 = 1.$$

Exactly one quotient dimension is revealed and one remains unresolved.

5.7 A distinction between revelation and identification

The revelation identity should not be interpreted as saying that every revealed quotient dimension is sufficient for exact inverse identification.

If

$$r_{\mathrm{rev}} < k,$$

then a positive-dimensional set of quotient states remains compatible with the observed outcome:

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}} > 0.$$

Under the regularity conditions above, this residual set corresponds to positive conditional information dimension.

Exact inverse identification therefore requires the stronger condition

$$d_I(Q \mid Y = y) = 0$$

together with the appropriate global uniqueness conditions. In the regular full-dimensional case, this corresponds locally to

$$r_{\mathrm{rev}} = k.$$

Even then, zero-dimensional projected fibers need not contain a single point globally. Discrete symmetries, multiple disconnected solutions, or other global nonuniqueness can leave finitely many compatible quotient states despite zero local dimension. Thus the information-dimension result is fundamentally a statement about the local dimensional structure of inverse uncertainty; global exact identification requires the structural and topological conditions developed earlier.

5.8 Interpretation

The preceding results establish the following chain:

$$\boxed{\text{latent disturbance} \longrightarrow \text{causal quotient} \longrightarrow \text{projected fiber} \longrightarrow \text{conditional information dimension.}}$$

The first step removes latent variation that has no causal consequence. The second determines which quotient states remain compatible with the observed outcome. The third measures the intrinsic dimension of that remaining uncertainty.

Under regularity,

$$\boxed{d_I(Q \mid Y) = k - r_{\mathrm{rev}}.}$$

Thus the quotient dimension is conserved through a revelation decomposition:

$$\boxed{\underbrace{d_I(Q)}_{\text{causal latent complexity}} = \underbrace{\mathcal{R}_{\mathrm{IC}}}_{\text{revealed information}} + \underbrace{d_I(Q \mid Y)}_{\text{unresolved inverse information}}.}$$

When the quotient is full-dimensional,

$$d_I(Q) = k,$$

the revelation dimension is exactly the differential revelation rank:

$$\boxed{\mathcal{R}_{\mathrm{IC}} = r_{\mathrm{rev}}.}$$

This identity is the information-theoretic counterpart of the projected-fiber dimension theorem. It shows that the local rank of the causal mechanism, after removing directions attributable to the causal antecedent X, determines how many dimensions of causally relevant latent information the outcome can reveal. The remaining dimensions constitute the intrinsic information burden of reversing the mechanism.

The next step is to convert this dimensional characterization into an operational measure of inverse difficulty. Rather than asking only how many dimensions remain unresolved, we ask how many bits of quotient information must be supplied, as a function of a desired reconstruction accuracy. This leads to the inverse-causal rate-distortion problem.

6 Canonical Examples and Comparative Implications

The preceding sections establish a general theory of causal information loss and inverse reconstruction. This section illustrates the theory through a collection of canonical mechanisms. The examples are chosen to separate four logically distinct phenomena: causal redundancy in the latent state, structural noninvertibility, observational revelation of the quotient, and the operational cost of reconstructing an unresolved quotient.

Throughout, the causal quotient is defined by the response-function map

$$Q = \Phi_f(U),$$

rather than by an arbitrary latent parametrization.

6.1 Additive latent disturbance: zero causal information loss, positive inverse burden

Consider the scalar mechanism

$$Y = h(X) + U,$$

where h is injective.

For a fixed u, the response function is

$$f_u(x) = h(x) + u.$$

Hence

$$\Phi_f(u) = f_u$$

is injective, so

$$Q \cong U.$$

There is therefore no causal information loss in passing from U to the quotient:

$$\sigma(Q) = \sigma(U).$$

The structural inverse is

$$X = h^{-1}(Y - Q).$$

Thus, conditional on Y, the unresolved quotient is precisely the latent disturbance.

Suppose U is a regular one-dimensional continuous variable. Then

$$d_I(Q) = 1.$$

If X and Q vary independently and $Y = h(X) + Q$, the quotient derivative has rank zero:

$$r_{\text{rev}} = 0.$$

Consequently,

$$d_I(Q \mid Y) = 1.$$

The inverse-causal rate therefore satisfies

$$R_{\text{IC}}(\varepsilon) = \frac{1}{2} \log \frac{1}{\varepsilon} + O(1).$$

This example establishes an important distinction. The latent variable is not causally redundant: the entire disturbance determines a distinct response function. Nevertheless, the outcome does not reveal any quotient dimension because the effect of Q is locally indistinguishable from an appropriate change in X.

Thus

$$\boxed{\text{CIL} = 0, \qquad \mathcal{R}_{\text{IC}} = 0, \qquad \mathfrak{d}_{\text{IC}} = 1.}$$

Causal information loss and inverse information burden are therefore fundamentally different quantities.

6.2 Squared disturbance: causal information loss without inverse ambiguity

Now consider

$$Y = X + U^2.$$

For fixed u,

$$f_u(x) = x + u^2.$$

Hence

$$f_u = f_{u'} \iff u^2 = (u')^2.$$

The states u and $-u$ are causally equivalent, and the canonical quotient is

$$\boxed{Q = U^2.}$$

The quotient map is therefore many-to-one:

$$U \longrightarrow Q.$$

For a discrete symmetric disturbance,

$$H(U \mid Q) > 0,$$

so causal information loss is strictly positive.

Yet the structural inverse is

$$X = Y - Q.$$

Once the quotient is available, no further ambiguity remains:

$$H(X \mid Y, Q) = 0.$$

This example therefore satisfies

$$\boxed{\mathrm{CIL} > 0, \qquad \eth_{\mathrm{IC}} = 0 \quad \text{given } Q.}$$

The information discarded by the causal quotient is not information needed to reverse the mechanism. The sign of U changes the latent state but does not change the response of Y to any intervention on X.

This illustrates why the quotient is the appropriate primitive for inverse causal analysis. Measuring the information content of U itself would incorrectly treat causally irrelevant distinctions as part of the inverse problem.

6.3 Mixed latent disturbance: quotienting beyond coordinatewise sufficiency

Consider

$$Y = X + U_1 + U_2^2.$$

A natural but noncanonical representation of the latent state is

$$(U_1, U_2^2).$$

However, the response function is

$$f_{u_1,u_2}(x) = x + u_1 + u_2^2.$$

Two latent states therefore generate the same response function whenever

$$u_1 + u_2^2 = u_1' + (u_2')^2.$$

The canonical quotient is consequently

$$\boxed{Q = U_1 + U_2^2.}$$

This example demonstrates that the quotient need not coincide with a subset of the original latent coordinates, nor with a coordinatewise sufficient statistic. The response-function map automatically identifies the combinations of latent variables that have identical causal consequences.

The inverse mechanism is

$$X = Y - Q.$$

Thus the relevant inverse uncertainty is uncertainty about the scalar combination Q, not uncertainty about the two-dimensional latent representation (U_1, U_2).

If (U_1, U_2) is continuously distributed and Q is regular, the inverse information dimension is one rather than two:

$$\eth_{\mathrm{IC}} = 1$$

before accounting for any revelation by Y.

The example therefore illustrates a second form of dimensional reduction:

$$\boxed{\dim(U) \ > \ \dim(Q).}$$

This reduction is generated by causal equivalence rather than by observation.

6.4 A noninvertible mechanism: latent information cannot repair structural ambiguity

Consider instead

$$Y = X^2 + Q,$$

where the support of X contains both x and $-x$ for some nonzero x.
Even if Q is observed exactly,

$$Y - Q = X^2$$

does not determine the sign of X. Hence there is no measurable function

$$g$$

such that

$$X = g(Y, Q)$$

almost surely under a distribution assigning positive probability to both branches.
The failure is structural. No additional information about the quotient can repair it.
This yields the fundamental distinction

$$\boxed{\text{latent-information deficiency} \neq \text{structural noninvertibility.}}$$

In the first case, the structural inverse exists but necessary quotient information is unavailable. In the second case, even the complete causal quotient is insufficient because the mechanism itself maps multiple values of X to the same outcome.

The distinction is central to the theory. Rate-distortion analysis quantifies the information required to overcome latent uncertainty only after the structural inversion problem has been well posed.

6.5 Partial revelation: an outcome can identify some causal dimensions but not others

Consider

$$X = (X_1, X_2), \qquad Q = (Q_1, Q_2),$$

and

$$Y = (X_1 + Q_1, Q_2).$$

The first quotient coordinate is confounded with X_1, while the second is directly revealed by the outcome.

Indeed,

$$D_x F = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad D_q F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The image of $D_x F$ is the first coordinate direction in outcome space. Modulo this image, only the second quotient direction remains:

$$r_{\mathrm{rev}} = 1.$$

Since

$$k = 2,$$

the projected fiber has dimension

$$\dim \mathcal{A}_y = 2 - 1 = 1.$$

Under the regular conditional distribution assumptions of Section 5,

$$d_I(Q \mid Y = y) = 1.$$

If the unconditional quotient is full-dimensional,

$$d_I(Q) = 2,$$

then

$$\boxed{d_I(Q) - d_I(Q \mid Y) = 1 = r_{\mathrm{rev}}.}$$

The inverse-causal rate therefore has leading behavior

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{1}{2} \log \frac{1}{\varepsilon} + O(1).}$$

One quotient dimension is already supplied by the observation; one must still be resolved for accurate inverse reconstruction.

6.6 Direct quotient revelation

At the opposite extreme, consider

$$Y = (X, Q).$$

The quotient is directly observed. Consequently,

$$P(Q \mid Y)$$

is degenerate and

$$d_I(Q \mid Y) = 0.$$

The quotient derivative has full rank:

$$r_{\mathrm{rev}} = k.$$

Thus

$$\dim \mathcal{A}_y = 0.$$

The leading inverse-causal rate vanishes:

$$\mathfrak{d}_{\mathrm{IC}} = 0.$$

This does not imply that zero bits are needed to communicate an exact value of a continuous Q in an ordinary coding system. Rather, it means that once Y is observed, no continuously varying quotient uncertainty remains. Any remaining ambiguity is at most zero-dimensional and therefore concerns global discrete identification rather than a logarithmically divergent continuous information burden.

6.7 A nonlinear mechanism with partial revelation

The framework is not restricted to additive mechanisms. Consider

$$Y = \begin{pmatrix} X_1 + Q_1^2 \\ X_2 + Q_2 \end{pmatrix}.$$

The response-function quotient is

$$Q^\star = (Q_1^2, Q_2).$$

Thus the sign of Q_1 is causally redundant.

The structural inverse is

$$X_1 = Y_1 - Q_1^2, \qquad X_2 = Y_2 - Q_2.$$

Consequently, the causally relevant latent state is two-dimensional even though the original disturbance may contain an additional discrete distinction.

Now suppose instead that the observed outcome is

$$Y = \begin{pmatrix} X_1 + Q_1^2 \\ X_2 + Q_2 \\ Q_2 \end{pmatrix}.$$

The second quotient coordinate is directly revealed. The first remains confounded with X_1.

The revelation rank is therefore one:

$$r_{\mathrm{rev}} = 1,$$

while the quotient dimension is two. Hence

$$d_I(Q \mid Y) = 1.$$

The example demonstrates that causal quotienting and observational revelation operate at different stages:

$$U \longrightarrow Q \longrightarrow Q \mid Y.$$

The first can remove discrete or continuous redundancy; the second can remove additional dimensions through observation.

6.8 A unifying comparison

The canonical examples can be summarized by the following decomposition:

$$\boxed{d_I(Q) = \underbrace{\mathcal{R}_{\mathrm{IC}}}_{\text{revealed}} + \underbrace{\mathfrak{d}_{\mathrm{IC}}}_{\text{unresolved}}.}$$

The distinct mechanisms correspond to different locations along this decomposition.

Mechanism	Causal quotient	Revelation	Inverse burden	height$Y = h(X) + U$
$Q = U$	0	k		
$Y = X + U^2$	$Q = U^2$	k		0
$Y = X + U_1 + U_2^2$	$Q = U_1 + U_2^2$	0		1
$Y = (X_1 + Q_1, Q_2)$	$Q = (Q_1, Q_2)$	1		1
$Y = (X, Q)$	Q	k		0

The table makes visible three distinct transformations that are sometimes conflated in inverse problems.

First, the raw latent state can contain causally irrelevant variation:

$$\dim(U) > \dim(Q).$$

Second, observation can reveal some but not all dimensions of Q:

$$0 < r_{\mathrm{rev}} < d_I(Q).$$

Third, the structural mechanism itself can fail to be invertible, in which case additional latent information cannot restore uniqueness.

These are mathematically distinct sources of inverse difficulty.

6.9 Implications for inverse causal analysis

The examples suggest a hierarchy of questions for an inverse causal problem.

Question 1: What latent distinctions matter causally? This is determined by

$$Q = \Phi_f(U).$$

Latent distinctions that do not change the intervention-response function should not enter the inverse information budget.

Question 2: Is the structural mechanism invertible in the causal antecedent? If no, exact inverse identification is impossible even with complete quotient information.

Question 3: How much of the causal quotient is revealed by observation? This is characterized locally by

$$r_{\mathrm{rev}} = \operatorname{rank} \overline{D}_q F.$$

Under the regular conditional framework,

$$d_I(Q \mid Y) = d_I(Q) - r_{\mathrm{rev}}.$$

Question 4: How much information must be supplied for reconstruction? This is answered operationally by

$$R_{\mathrm{IC}}(\varepsilon).$$

In the regular finite-dimensional regime,

$$R_{\mathrm{IC}}(\varepsilon) = \frac{d_I(Q \mid Y)}{2} \log \frac{1}{\varepsilon} + O(1).$$

The resulting hierarchy can therefore be written as

$$\boxed{\begin{array}{c}\text{causal equivalence}\\ \downarrow \\ \text{causal quotient } Q \\ \downarrow \\ \text{structural invertibility} \\ \downarrow \\ \text{projected-fiber dimension} \\ \downarrow \\ \text{conditional information dimension} \\ \downarrow \\ \text{inverse-causal rate.}\end{array}}$$

Each level answers a different question. None can, in general, be substituted for another.

6.10 The central decomposition

The examples support the paper's central interpretation of inverse causal inference:

$$\boxed{\text{inverse difficulty} = \text{causal latent complexity} - \text{causal revelation},}$$

with the terms made precise, in the regular full-dimensional setting, as

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = d_I(Q) - r_{\mathrm{rev}}.}$$

The first term is determined by the causal response-function quotient. The second is determined by the extent to which the observed outcome separates causally distinct quotient states after accounting for variation in the antecedent.

This decomposition also clarifies why observational uncertainty alone is not the fundamental object. The relevant uncertainty is uncertainty over *causally distinct mechanisms*. Two latent states that are observationally distinguishable but generate the same response function should not count as distinct causal states. Conversely, two causally distinct states may remain observationally indistinguishable after conditioning on the outcome.

The theory therefore separates three notions that are often collapsed:

$$\boxed{\text{latent distinction} \quad \neq \quad \text{causal distinction} \quad \neq \quad \text{observational revelation.}}$$

The causal quotient isolates the second. The projected fiber measures the third. Rate-distortion translates the residual gap into an operational reconstruction cost.

6.11 Transition to the general theory

The examples also reveal the scope of the theory. The leading-order results depend on regular finite-dimensional geometry, but the underlying construction does not. The response-function quotient remains meaningful in much greater generality, while the projected-fiber dimension and information-dimension results require increasingly stronger regularity assumptions.

This suggests a natural hierarchy of the theory:

Level I:	causal quotient and structural equivalence,
Level II:	exact inverse identification,
Level III:	smooth projected-fiber geometry,
Level IV:	conditional information dimension,
Level V:	inverse-causal rate-distortion.

The lower levels require relatively weak assumptions; the higher levels provide progressively sharper quantitative conclusions.

This hierarchy is important for applications. A researcher may be able to determine the causal quotient without having a smooth model, or establish structural invertibility without possessing a density suitable for information-dimension analysis. The framework does not require every problem to reach the final rate-distortion layer.

The central object remains the same throughout:

$$\boxed{Q = \Phi_f(U).}$$

What changes is the mathematical structure imposed on the quotient and on its relationship to the observed outcome.

The next section develops the broader implications of this hierarchy for causal identification, inverse problems, and the measurement of causal information. It also distinguishes the present notion of causal information loss from existing notions of statistical sufficiency, information flow, and representation equivalence.

7 Discussion and Relation to Existing Concepts

The preceding results establish a unified theory of causal information loss, inverse identification, geometric revelation, and inverse reconstruction. This section places the theory relative to several neighboring concepts. The purpose is not to replace existing notions of sufficiency, causal information, inverse problems, or representation equivalence, but to clarify the precise object introduced here and the questions for which it is designed.

The central distinction is between three levels of information:

$$\boxed{\text{latent information} \quad \longrightarrow \quad \text{causal information} \quad \longrightarrow \quad \text{observationally revealed information.}}$$

The response-function quotient defines the second level. Conditional probability and information dimension characterize the third.

7.1 Relation to causal identification

Causal identification asks whether a causal quantity is uniquely determined by the available distributional or experimental information. The present theory addresses a related but distinct question:

given a structural mechanism, how much information about its causally relevant latent state is required to reverse that mechanism?

This distinction matters because exact inverse identification has two conceptually different failure modes.

First, the mechanism itself may be noninvertible:

$$Y = F(X, Q)$$

may map multiple values of X to the same (Y, Q). In that case, no additional information about the latent quotient can restore exact recovery.

Second, the mechanism may be structurally invertible once Q is known:

$$X = g(Y, Q),$$

but Q may not be available to the inverse analyst.

The theory therefore decomposes inverse identification into

$$\boxed{\text{structural invertibility} + \text{quotient availability.}}$$

This is different from a conventional identification statement of the form

$$X = \varphi(P)$$

for some functional φ of an observed distribution P. The present framework instead studies the information content of the structural state that must be supplied to make the inverse map well defined.

The distinction is particularly relevant when the forward causal model is itself known. Knowing the forward mechanism does not imply that its inverse is identified from the outcome alone. The missing object may be latent state information rather than uncertainty about the mechanism.

7.2 Relation to inverse problems

Classical inverse-problem theory distinguishes existence, uniqueness, and stability of solutions. The present theory complements this framework by identifying an information-theoretic layer beneath inverse reconstruction.

Consider

$$Y = F(X, Q).$$

The standard inverse-problem question asks whether X can be recovered from Y. The present framework additionally asks:

1. which components of the latent state Q are actually causally relevant;
2. which dimensions of Q remain compatible with the observed Y;
3. how much information about those unresolved dimensions is required for reconstruction.

The corresponding mathematical objects are

$$Q = \Phi_f(U),$$

$$\mathcal{A}_y = \{q : \exists x,\ F(x, q) = y\},$$

and

$$R_{\mathrm{IC}}(\varepsilon).$$

Thus the theory adds an information geometry to the usual existence–uniqueness–stability decomposition:

$$\boxed{\text{existence} \ \to \ \text{uniqueness} \ \to \ \text{stability} \ \to \ \text{information cost.}}$$

The distinction is substantive. A problem can be uniquely solvable but require an arbitrarily large amount of information for high-precision reconstruction, or it can fail uniqueness because of a positive-dimensional projected fiber. Conversely, a zero-dimensional projected fiber can still contain several globally distinct solutions. Geometric uniqueness and information cost therefore need not coincide.

7.3 Relation to sufficient statistics

The causal quotient resembles a sufficient representation, but the analogy must be used carefully.

A statistical sufficient statistic preserves information about a specified parameter relative to a statistical experiment. The quotient

$$Q = \Phi_f(U)$$

is instead defined by an equivalence relation on latent states:

$$u \sim_f u' \quad \Longleftrightarrow \quad f(\cdot, u) = f(\cdot, u').$$

Its defining criterion is therefore equality of the complete response function to interventions on X, not preservation of a likelihood or conditional distribution.

This difference has several consequences.

First, the quotient is tied to a structural causal model rather than to a particular statistical sample.

Second, it is defined relative to a specified intervention domain. If the set of interventions on X changes, the relevant response-function equivalence relation may also change.

Third, the quotient can be strictly smaller than the latent state even when every component of the latent state is statistically observable. What matters is not whether a latent distinction can be measured, but whether that distinction changes the causal response function.

For these reasons, we use the term *causal quotient* rather than simply *sufficient statistic*. The latter would obscure the structural criterion that defines the equivalence classes.

7.4 Relation to latent-variable representations

Latent-variable models typically admit multiple parameterizations of the same observable structure. The response-function quotient provides a particularly direct way to remove one source of this representational redundancy.

Suppose

$$U \mapsto \widetilde{U}$$

is a reparameterization under which

$$f(\cdot, U) = \widetilde{f}(\cdot, \widetilde{U}).$$

Different parameter values may therefore describe the same causal response function.

The quotient construction identifies all such states:

$$U \longrightarrow Q.$$

This is important for the inverse information measure. If inverse complexity were measured directly in the coordinates of U, it could depend on arbitrary latent parameterization. By contrast, the quotient retains only distinctions that alter causal behavior.

The information-dimension and rate-distortion formulations then provide a second layer of representation invariance. Under bi-Lipschitz changes of quotient coordinates, the relevant information dimension and leading rate exponent are preserved.

The resulting hierarchy is

$$\boxed{\text{latent representation} \ \rightarrow\ \text{causal equivalence class} \ \rightarrow\ \text{intrinsic inverse complexity.}}$$

7.5 Relation to causal information flow

Information-theoretic approaches to causality have developed measures of information flow, directed information, transfer entropy, and related quantities. Those concepts generally quantify how statistical information about one variable is transmitted through or associated with another variable.

The present object is different.

The causal revelation quantity

$$\mathcal{R}_{\mathrm{IC}} = d_I(Q) - d_I(Q \mid Y)$$

asks how much of a *causally relevant latent state* becomes identifiable from an observed outcome. Its causal content therefore enters before the information-theoretic calculation:

$$U \xrightarrow{\Phi_f} Q.$$

Only after causally redundant latent states have been collapsed do we measure

$$Q \mid Y.$$

This ordering is essential. A large amount of statistical information about U need not imply a large amount of causal information about the mechanism, because some distinctions in U may be causally irrelevant. Conversely, a small observational signal can be highly informative about a low-dimensional causal quotient.

The theory therefore concerns *information about causal state*, rather than information flow between variables in the usual sense.

7.6 Relation to causal representation learning

Causal representation learning seeks representations that preserve causal structure, often under assumptions concerning disentanglement, invariance, interventions, or latent-variable structure. The present framework shares the objective of distinguishing causally meaningful latent structure from arbitrary representation.

The difference is that the causal quotient is defined by a stronger functional criterion:

$$u \sim_f u' \quad \Longleftrightarrow \quad f(x,u) = f(x,u') \quad \forall x.$$

Rather than asking whether a representation is useful for causal prediction or discovery, the quotient asks whether two latent states are indistinguishable with respect to their entire response functions.

This gives a natural target for causal representation learning:

$$\boxed{\text{learn a representation of the causal quotient.}}$$

The present paper does not propose an estimator for this object. Instead, it characterizes what follows if the quotient is known: structural reversibility, projected-fiber geometry, information dimension, and rate-distortion.

This distinction is deliberate. Learning the quotient is an empirical problem; characterizing its information geometry is a theoretical one.

7.7 Relation to counterfactual and abductive reasoning

There is also a close conceptual relationship to abductive and counterfactual reasoning in structural causal models.

Given an observed outcome, one may infer latent disturbances or structural states and then use those states to answer counterfactual questions. The present framework formalizes a complementary issue: when is the latent state itself recoverable, and how much causally relevant latent information is required to recover it?

The quotient is especially useful in this setting because counterfactual predictions need not depend on every detail of the latent disturbance. If

$$f(\cdot, u) = f(\cdot, u'),$$

then u and u' generate identical responses to every intervention on X and are therefore interchangeable for all counterfactual questions involving that intervention domain.

Hence the quotient can be interpreted as the state of latent information that remains relevant for the entire family of such interventions.

7.8 Local revelation versus global identification

One of the most important limitations of the geometric results is that

$$r_{\text{rev}}$$

is a local quantity.

The quotient derivative determines how many infinitesimal quotient directions are distinguishable after accounting for perturbations in X. It does not by itself determine the global number of compatible quotient states.

For example, a smooth mechanism can have

$$r_{\text{rev}} = k$$

everywhere while the global projected fiber contains several isolated points. In that case,

$$\dim \mathcal{A}_y = 0,$$

but global inverse identification can still fail.

The theory consequently distinguishes

$$\boxed{\text{local dimensional revelation} \neq \text{global uniqueness.}}$$

The former is captured by differential geometry and information dimension. The latter requires additional global restrictions such as injectivity, connectedness, monotonicity, or suitable topological conditions.

This distinction is not merely technical. Information dimension deliberately ignores finite discrete multiplicities. A finite number of isolated possibilities has zero information dimension even though distinguishing among them may require a nonzero number of bits.

Thus the theory separates continuous inverse uncertainty from discrete global ambiguity.

7.9 Causal information loss versus inverse information burden

Two quantities introduced in the paper deserve particular separation.

For discrete U,

$$\mathrm{CIL} = H(U \mid Q)$$

measures information discarded when the latent disturbance is replaced by its causal quotient.

By contrast,

$$R_{\mathrm{IC}}(\varepsilon)$$

measures the information that must be supplied about the quotient for approximate inverse reconstruction.

These quantities answer opposite questions.

Causal information loss asks:

$$\boxed{\text{What information in the latent state is causally redundant?}}$$

Inverse rate-distortion asks:

$$\boxed{\text{What causally relevant information is still missing after observing } Y\text{?}}$$

The two need not move together.

For

$$Y = X + U^2,$$

the sign of U is discarded:

$$H(U \mid U^2) > 0,$$

but the discarded information is irrelevant for recovering X.

For

$$Y = h(X) + U,$$

there may be no causal information loss at all:

$$Q = U,$$

yet the quotient remains completely unresolved by Y in the dimensional sense when X and U are continuously varying.

This gives the conceptual distinction

$$\boxed{\text{information lost by causal compression} \neq \text{information missing for causal inversion.}}$$

7.10 Causal revelation as a quotient-relative concept

The revelation quantity

$$\mathcal{R}_{\mathrm{IC}} = d_I(Q) - d_I(Q \mid Y)$$

is also inherently relative to the causal intervention domain used to define Q.

If the intervention domain expands, two latent states that previously generated identical responses may become distinguishable. The quotient can therefore become finer:

$$Q_{\mathcal{X}_1} \quad \longrightarrow \quad Q_{\mathcal{X}_2}, \qquad \mathcal{X}_1 \subseteq \mathcal{X}_2.$$

This observation emphasizes that there is no intervention-free notion of causal latent information in the present framework. Causal relevance is always defined relative to the family of interventions whose response functions are being considered.

This is analogous to the fact that identification itself depends on the target. A latent distinction can be irrelevant for one causal question and relevant for another.

7.11 Relation to statistical sufficiency of the outcome

The framework also clarifies a potential misconception concerning the outcome Y.

If

$$P(X \mid Y)$$

is highly concentrated, this does not necessarily imply that Y reveals the causal quotient. Conversely, Y may reveal a large portion of the quotient while remaining insufficient to identify X because of structural noninvertibility.

The relevant object is not simply the predictive quality of Y for X. Instead, the question is whether variation in Q produces outcome variation that cannot be replicated by variation in X.

This is exactly what the quotient derivative captures:

$$\overline{D}_q F : T_q\mathcal{Q} \to T_y\mathcal{Y}/\operatorname{Im} D_x F.$$

A quotient direction is locally revealed precisely when its effect on Y survives projection onto the space of outcome variations that can already be generated by changing X.

Thus the theory gives a causal interpretation to an otherwise purely geometric quotient construction.

7.12 What is genuinely new in the framework

The individual mathematical ingredients used in the paper are established objects. Response functions, structural causal models, disintegration, constant-rank geometry, Hausdorff measure, information dimension, and rate-distortion theory each have substantial prior literatures.

The contribution lies in their organization around a common causal state space.

The central construction is

$$Q = \Phi_f(U),$$

the quotient of latent states by equality of their intervention-response functions. From this object, the paper derives a sequence of results:

$$\boxed{\begin{array}{c} Q \longrightarrow \text{inverse identification} \longrightarrow \text{projected-fiber geometry} \\ \longrightarrow \text{conditional information dimension} \longrightarrow \text{inverse-causal rate.} \end{array}}$$

The projected-fiber result gives

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}},$$

the conditional information-dimension result gives

$$d_I(Q \mid Y) = k - r_{\mathrm{rev}},$$

under regularity, and the rate-distortion result gives

$$R_{\mathrm{IC}}(\varepsilon) = \frac{k - r_{\mathrm{rev}}}{2} \log \frac{1}{\varepsilon} + O(1)$$

in the corresponding finite-dimensional regime.

The resulting chain is therefore not simply an application of information theory to a causal model. It is a theory in which the causal model determines the relevant state space, geometry determines the unresolved dimensions of that state space, and information theory determines the operational cost of resolving them.

7.13 Scope and limitations

The theory also has clear boundaries.

First, the causal quotient depends on the intervention domain. It is not an absolute property of a latent variable independent of the causal question.

Second, the projected-fiber and information-dimension results rely on regularity assumptions. Singular conditional distributions, varying rank, fractal supports, and mixtures of components with different dimensions require separate treatment.

Third, the rate-distortion results characterize asymptotic continuous information cost. They do not by themselves account for finite discrete ambiguities, which are invisible to information dimension.

Fourth, the framework assumes that the relevant structural mechanism is known or otherwise sufficiently characterized. Learning the response-function quotient from observational or experimental data is a separate statistical problem.

These limitations are also opportunities. In particular, extending the theory to singular quotient measures, stratified fibers, and partially unknown structural mechanisms would connect causal identification more directly to modern inverse problems and nonparametric representation learning.

7.14 Broader interpretation

The theory suggests a general way to think about causal information.

A latent state contains potentially many distinctions. Some change the causal response of the system; others do not. The response-function quotient removes the latter. Observation then reveals some portion of the remaining causal state. What remains is inverse uncertainty.

The conceptual decomposition is

$$\boxed{\begin{aligned} \text{raw latent complexity} &\longrightarrow \text{causal latent complexity}, \\ \text{causal latent complexity} &\longrightarrow \text{revealed causal information} + \text{unresolved causal information}. \end{aligned}}$$

In the regular full-dimensional setting,

$$\boxed{d_I(Q) = \underbrace{r_{\mathrm{rev}}}_{\text{revealed}} + \underbrace{d_I(Q \mid Y)}_{\text{unresolved}},}$$

and the unresolved component has operational cost

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{d_I(Q \mid Y)}{2} \log \frac{1}{\varepsilon} + O(1).}$$

This provides a precise interpretation of the inverse problem as an information problem. Structural causality determines which latent distinctions matter; observation determines which of those distinctions survive; and reconstruction requires communicating what observation has not revealed.

The principal conceptual contribution is therefore not a new measure of generic statistical dependence. It is a causal decomposition of latent information into *causally redundant*, *observationally revealed*, and *inverse-relevant* components.

That decomposition provides a common language for structural causal models, inverse problems, differential geometry, information dimension, and rate-distortion theory.

8 Extensions and Open Problems

The theory developed above is intentionally formulated around a regular finite-dimensional regime in which the causal quotient is measurable, the structural mechanism is smooth, projected fibers have constant dimension, conditional quotient laws are geometrically regular, and the inverse map admits a high-resolution rate-distortion characterization. These assumptions permit a sharp chain of results from causal equivalence to geometric revelation and operational inverse complexity.

At the same time, the framework points toward a broader theory. This section identifies several directions in which the present results can be extended. Some concern mathematical generality; others concern identification and estimation when the structural mechanism is not known.

8.1 Singular and fractal causal quotients

The finite-dimensional manifold setting is not the only regime in which causal quotient states can arise. A quotient may have a singular distribution, a noninteger information dimension, or support on a fractal set.

The rate-distortion formulation already accommodates this possibility at a conceptual level. If

$$d_I(Q \mid Y)$$

exists but is not an integer, the high-resolution inverse-causal rate is expected to satisfy

$$R_{\mathrm{IC}}(\varepsilon) \sim \frac{d_I(Q \mid Y)}{2} \log \frac{1}{\varepsilon}.$$

A general theory would therefore replace smooth manifold dimension with information dimension throughout. The geometric object would no longer necessarily be an embedded manifold, while the operational object would remain well defined.

A particularly interesting question is whether the causal quotient can generate genuinely noninteger inverse dimensions:

$$d_I(Q \mid Y) \notin \mathbb{N}.$$

Such cases would provide causal examples in which inverse complexity is intrinsically fractional rather than merely a consequence of a high-dimensional parametrization.

This suggests a broader notion of

$$\boxed{\text{causal information dimension}}$$

that does not presuppose smooth geometry.

8.2 Rank-varying mechanisms and stratified projected fibers

The constant-rank assumption provides the clean identity

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}}.$$

In general, however,

$$r_{\mathrm{rev}}(x, q)$$

may vary across the structural support.

The projected fiber may then decompose into strata

$$\mathcal{A}_y = \bigcup_j \mathcal{A}_{y,j},$$

with

$$\dim \mathcal{A}_{y,j} = k - r_{\mathrm{rev},j}.$$

The conditional quotient distribution can consequently place mass on components with different dimensions. The resulting information dimension need not equal a simple average of local revelation ranks.

A general rank-varying theory would need to characterize which strata dominate the small-scale entropy and rate-distortion asymptotics. This raises a natural question:

$$\boxed{\text{What is the correct probabilistic aggregation of local causal revelation ranks?}}$$

One possibility is that the maximal-dimensional strata carrying asymptotically nonnegligible probability determine the leading information dimension. More delicate mixtures could require a genuinely multifractal treatment.

8.3 Discrete global ambiguity

The information-dimension framework deliberately captures continuously varying uncertainty. It does not measure finite discrete ambiguity.

Suppose

$$\mathcal{A}_y = \{q_1, \ldots, q_J\}$$

with $J > 1$. Then

$$\dim \mathcal{A}_y = 0$$

and

$$d_I(Q \mid Y = y) = 0,$$

even though distinguishing the J possibilities can require approximately

$$\log_2 J$$

bits.

This suggests decomposing inverse complexity into continuous and discrete components:

$$\boxed{\text{inverse complexity} = \text{continuous information dimension} + \text{discrete identification complexity.}}$$

A complete theory could therefore combine information dimension with conditional discrete entropy, or more generally with a decomposition of conditional measures into atomic and nonatomic components.

Such a decomposition would extend the present framework from local dimensional ambiguity to full global inverse identification.

8.4 Stability and ill-posed inverse mechanisms

The current rate-distortion results quantify information required for reconstruction, but they do not fully characterize numerical stability.

Two mechanisms may have identical inverse information dimension

$$\mathfrak{d}_{\mathrm{IC}} = k$$

while differing dramatically in conditioning. If

$$J_q g(y, q)$$

is nearly singular in some region, a small error in quotient information can generate a large error in X.

The local metric

$$G(y, q) = J_q g(y, q)^\top J_q g(y, q)$$

already points toward this distinction. Its eigenvalues characterize the local amplification of quotient errors.

A natural extension is therefore to define an operational measure of inverse instability alongside inverse information dimension. One possibility would be to study the asymptotic rate under a distortion metric induced by the inverse Jacobian:

$$d_X(g(y, q), g(y, q'))$$

rather than Euclidean quotient distortion.

This would separate two properties of an inverse mechanism:

$$\boxed{\text{information complexity} \quad \text{versus} \quad \text{conditioning.}}$$

The first is captured by the leading rate exponent; the second enters the finite-resolution constant and potentially the behavior outside the asymptotic regime.

8.5 Unknown structural mechanisms

The present theory takes the response-function map

$$\Phi_f$$

as given. In empirical applications, however, f is generally unknown.

This creates a second inverse problem:

$$\boxed{\text{learning the causal quotient}}$$

before one can use the quotient for inverse reconstruction.

Suppose an estimator $\widehat{f}$ is obtained from experimental or observational data. It induces an estimated response-function map

$$\widehat{\Phi}_f$$

and therefore an estimated quotient

$$\widehat{Q}.$$

A fundamental question is whether

$$\widehat{Q} \to Q$$

in an appropriate sense as the amount of causal data increases.

This is substantially more difficult than ordinary prediction. Two estimated response functions may be close under an observational distribution while differing substantially on interventions outside the observed support. A statistical theory of causal quotient estimation would therefore require a topology or metric on response functions that reflects the intended intervention domain.

Potential questions include:

When is the causal quotient statistically identifiable?
How many interventions are required to learn it?
How does intervention design affect quotient estimation?
How does estimation error propagate into $R_{\mathrm{IC}}(\varepsilon)$?

This connects the present theory to experimental design and active causal learning.

8.6 Unknown or partially known intervention domains

The quotient

$$Q = \Phi_f(U)$$

is defined relative to the family of interventions over which response functions are compared. If $\mathcal{X}_1$ is a restricted intervention domain and $\mathcal{X}_2$ is larger, then

$$u \sim_{\mathcal{X}_2} u' \quad \Longrightarrow \quad u \sim_{\mathcal{X}_1} u',$$

but the converse need not hold.

Hence the causal quotient can depend on the intervention domain:

$$Q_{\mathcal{X}_1} \neq Q_{\mathcal{X}_2}.$$

This raises a design question that is largely absent from the current formulation:

$$\boxed{\text{Which interventions maximize the causal resolution of latent states?}}$$

The quotient framework suggests a natural objective. One could choose interventions to maximize

$$r_{\mathrm{rev}}$$

or, more globally, to minimize

$$d_I(Q \mid Y).$$

This would turn causal experimental design into an information-geometric problem:

$$\boxed{\text{choose interventions that maximally reveal causally distinct latent states.}}$$

8.7 Dynamic and sequential mechanisms

The present framework is static:

$$Y = F(X, Q).$$

Many causal systems are dynamic,

$$Y_t = F_t(X_{0:t}, Q_t),$$

with latent states evolving over time.

In such settings, the quotient may itself be a process:

$$Q_{0:T}.$$

The inverse problem becomes sequential, and information revealed at one time can alter the information burden at later times.

A dynamic theory would naturally replace

$$I(Q; M \mid Y)$$

with a sequential or directed information quantity and study a dynamic inverse-causal rate:

$$R^{(T)}_{\mathrm{IC}}(\varepsilon).$$

A central open question is whether the static decomposition extends to an information rate:

$$\boxed{\text{causal latent information rate} = \text{revealed information rate} + \text{unresolved inverse information rate.}}$$

Such an extension could connect the framework to state-space models, filtering, control, and sequential causal inference.

8.8 Multiple latent mechanisms and causal networks

The presentation has focused on a single structural mechanism

$$Y = F(X, Q).$$

A broader structural causal model may contain a network of mechanisms,

$$X_1 = f_1(\mathrm{pa}_1, U_1), \quad \dots, \quad X_n = f_n(\mathrm{pa}_n, U_n).$$

In a network, causal equivalence may be defined locally at each mechanism or globally across the entire system. The global quotient could potentially be constructed from a collection of local response-function quotients.

This raises a compositional question:

$$\boxed{\text{How do causal information losses compose across a causal graph?}}$$

For independent mechanisms one might hope for an additive decomposition. With shared disturbances, feedback, or nonlinear interactions, however, quotient dimensions may not add straightforwardly.

A compositional theory could potentially provide a causal analogue of modular information accounting:

$$\text{global inverse complexity} = \text{local quotient complexities} - \text{cross-mechanism revelation.}$$

8.9 Partial identification

Exact inverse identification is only one endpoint. In many applications the goal is to characterize a set of admissible antecedents rather than recover a unique value.

The projected fiber already provides such a set:

$$\mathcal{A}_y = \{q : \exists x,\ F(x,q) = y\}.$$

A natural extension is to study functionals of this set:

$$\mathrm{diam}(\mathcal{A}_y), \qquad \mathcal{H}^d(\mathcal{A}_y), \qquad P(Q \in \mathcal{A} \mid Y = y),$$

or corresponding uncertainty regions for X.

This would connect the framework to partial identification. Instead of asking whether

$$X = g(Y, Q)$$

exists, one could characterize the geometry and information content of the identified set.

The resulting theory would distinguish:

$$\boxed{\text{point identification} \quad \leftrightarrow \quad \text{set identification} \quad \leftrightarrow \quad \text{information-limited reconstruction.}}$$

8.10 Beyond Euclidean distortion

The rate-distortion formulation currently uses squared-error reconstruction:

$$\mathbb{E}\|X - \widehat{X}\|^2 \le \varepsilon.$$

Different causal applications may require asymmetric or task-specific loss. For example, the cost of confusing two structural states may depend on the intervention under consideration.

One could therefore replace Euclidean distortion by a causal task loss

$$d_{\text{task}}(x, \widehat{x})$$

or even a loss defined directly on response functions:

$$d_{\text{resp}}\big(f_x, f_{\widehat{x}}\big).$$

The corresponding rate

$$R^d_{\text{IC}}(\varepsilon)$$

would measure the information required not merely to reconstruct X numerically, but to reproduce its causal consequences.

This raises a potentially important distinction:

$$\boxed{\text{state reconstruction} \neq \text{causal reconstruction.}}$$

In some applications, exact recovery of X may be unnecessary if different values of X produce nearly identical consequences for the policy interventions of interest.

8.11 Causal quotient learning and representation learning

The quotient construction also suggests an empirical learning problem.

Given interventional data

$$\{(x_i, y_i)\}_{i=1}^n$$

under varying latent states, one could seek a representation z_i such that

$$z_i = z_j$$

whenever the corresponding latent states generate the same response function, while distinct response functions remain separated.

An ideal representation would therefore approximate

$$Q = \Phi_f(U).$$

This provides a precise target for a class of causal representation-learning procedures. Rather than defining a representation through reconstruction accuracy alone, one could define it through invariance of the entire intervention-response function.

The theoretical problem is then to establish guarantees of the form

$$d(\widehat{Q}, Q) \to 0$$

and determine how representation error affects

$$\widehat{r}_{\mathrm{rev}}, \qquad \widehat{d}_I(Q \mid Y), \qquad \widehat{R}_{\mathrm{IC}}(\varepsilon).$$

This would provide a direct empirical counterpart to the theoretical framework developed here.

8.12 Causal information bottlenecks

The quotient construction may also provide a principled formulation of a causal information bottleneck.

Suppose one seeks a representation Z satisfying two objectives:

$$Z \approx Q$$

in the sense of preserving causal response functions, while minimizing

$$I(U; Z).$$

The trivial solution $Z = Q$ is causally complete, but may contain more information about the raw latent state than necessary for a particular task. One could therefore study

$$\inf I(U; Z)$$

subject to a prescribed causal fidelity criterion.

Such a formulation would differ from a standard information bottleneck because the retained information is constrained by causal equivalence rather than predictive sufficiency alone.

This suggests a broader optimization problem:

$$\boxed{\text{compress latent information while preserving causal response structure.}}$$

The present paper identifies the exact zero-loss quotient toward which such compression should converge.

8.13 Causal information budgets

The inverse-causal rate also suggests an information-budget interpretation.

Suppose an analyst has access to an auxiliary variable Z. Define

$$\mathrm{ICID}(Z) = H(X \mid Y, Z)$$

in the discrete setting, or its appropriate continuous analogue through reconstruction rate.

Then additional information can be evaluated according to how much it reduces inverse uncertainty:

$$\Delta_Z = R_{\mathrm{IC}}^{Y}(\varepsilon) - R_{\mathrm{IC}}^{Y,Z}(\varepsilon).$$

This creates a natural framework for comparing data sources by their value for inverse causal reconstruction.

For example, one could ask whether an additional measurement, intervention, instrument, or sensor reduces the inverse-causal rate more effectively than another. The relevant criterion is not simply predictive accuracy for Y, but reduction in the information burden required to recover the causal antecedent.

This suggests an information-theoretic approach to measurement design:

$$\boxed{\text{choose additional observations that most reduce inverse-causal information cost.}}$$

8.14 A broader mathematical program

Taken together, these extensions suggest a hierarchy of increasingly general causal information geometry:

$$\boxed{\begin{aligned} \text{smooth finite-dimensional quotients} &\longrightarrow \text{singular and fractal quotients}, \\ \text{constant rank} &\longrightarrow \text{stratified rank-varying mechanisms}, \\ \text{static mechanisms} &\longrightarrow \text{dynamic causal systems}, \\ \text{known mechanisms} &\longrightarrow \text{learned mechanisms}, \\ \text{state reconstruction} &\longrightarrow \text{causal-response reconstruction}. \end{aligned}}$$

The present paper establishes the first layer of this program. Its main purpose is to identify the objects that should remain invariant as the assumptions are relaxed:

$$\boxed{\text{causal quotient} \ \rightarrow \ \text{projected compatibility set} \ \rightarrow \ \text{conditional causal information} \ \rightarrow \ \text{operational inverse cost}.}$$

The mathematical details of the extensions may differ substantially, but the underlying logic remains the same.

8.15 Open questions

Several questions appear particularly fundamental.

1. Can the conditional information-dimension theorem be extended to singular and fractal projected fibers?
2. What is the correct revelation functional when r_{rev} varies across strata or when the conditional quotient law is a mixture of dimensions?

3. Can continuous inverse uncertainty and discrete global ambiguity be combined into a single operational measure of inverse causal complexity?

4. Under what conditions can the causal quotient itself be consistently estimated from finite interventional data?

5. How should intervention design be optimized to maximize quotient revelation?

6. Can causal information loss and inverse information burden be given compositional laws for general structural causal graphs?

7. Can the inverse-causal rate be generalized from state reconstruction to reconstruction of causal response functions or policy-relevant functionals?

8. What statistical procedures can estimate

$$r_{\text{rev}}, \qquad d_I(Q \mid Y), \qquad R_{\text{IC}}(\varepsilon)$$

from finite samples?

These questions point beyond the particular regularity regime analyzed here, but they do not require changing the central object. The response-function quotient remains the natural candidate for the causal state space on which inverse information should be measured.

8.16 Closing perspective

The theory developed in this paper can therefore be viewed as a starting point for a broader *causal information geometry.* The central premise is that causal systems transform information in a structured way: latent states are first identified up to their intervention-response functions, observations reveal some of the resulting causal state, and inverse reconstruction must recover what remains unresolved.

In the regular setting developed here, this logic produces the chain

$$\boxed{Q = \Phi_f(U) \quad \longrightarrow \quad \dim \mathcal{A}_y = k - r_{\text{rev}} \quad \longrightarrow \quad d_I(Q \mid Y) = k - r_{\text{rev}} \quad \longrightarrow \quad R_{\text{IC}}(\varepsilon) = \frac{k - r_{\text{rev}}}{2} \log \frac{1}{\varepsilon} + O(1).}$$

The extensions above ask how much of this chain survives when smoothness, finite dimensionality, constant rank, known mechanisms, and exact state reconstruction are abandoned.

That question is deliberately left open. The contribution of the present paper is to establish the conceptual and mathematical foundation on which those generalizations can be built.

9 Conclusion

This paper develops a theory of causal information loss and the information required to reverse a causal mechanism. The starting point is a simple observation: a structural causal model does not merely map causes into outcomes. It also determines which distinctions in the latent state remain causally relevant after the mechanism has operated. Once this distinction is made explicit, inverse causal inference can be formulated as a problem of recovering a causal state from an observed outcome together with whatever portion of the latent causal structure remains available.

The central object is the *causal quotient*

$$Q = \Phi_f(U), \qquad \Phi_f(u) = f_u,$$

which identifies latent states whenever they induce the same response function over the intervention domain. This quotient separates causal information from representational redundancy: two latent states that generate identical responses to every admissible intervention are indistinguishable for the causal mechanism and need not be separately recovered. Conversely, distinctions that survive in the response-function map are causally meaningful for reversal. The resulting quotient therefore provides a canonical state space on which the inverse problem can be posed.

This perspective yields a decomposition of inverse identification into two conceptually distinct requirements. First, the structural mechanism must be invertible in the causal variable. If distinct values of the causal antecedent generate the same outcome for a fixed causal latent state, no amount of information about that latent state can restore exact inversion. Second, the information needed to distinguish the relevant causal latent states must be available. Under appropriate measurability conditions, quotient-complete information permits the structural decoder to recover the antecedent. Thus, failure of inverse identification can arise either from noninvertibility of the mechanism or from information loss about the causal quotient, and these two failures should not be conflated.

The geometric results make the second component quantitative. For a smooth mechanism

$$Y = F(X, Q),$$

the relevant object is not the rank of the derivative with respect to Q alone, because changes in Q that can be offset by changes in X are not observationally distinguishable. Instead, the derivative of the quotient state must be considered modulo the outcome directions generated by X. The resulting revelation rank r_{rev} measures the number of local causal-quotient directions revealed by the observation. Under the regularity conditions of the projected-fiber theorem,

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}},$$

where $\mathcal{A}_y$ is the set of quotient states compatible with the observed outcome. The unresolved projected fiber therefore provides a geometric representation of inverse causal ambiguity.

The information-theoretic results provide a complementary quantitative characterization. When the conditional quotient law is regular on the projected fiber,

$$d_I(Q \mid Y = y) = k - r_{\mathrm{rev}}$$

for almost every outcome. Consequently, when the unconditional quotient has full information dimension k,

$$\boxed{d_I(Q) = r_{\mathrm{rev}} + d_I(Q \mid Y).}$$

The causal complexity of the latent state is thus decomposed into the part revealed by the outcome and the part that remains unresolved. This identity connects the geometry of projected fibers to the information dimension of the remaining inverse problem.

The same quantity governs the asymptotic information required for approximate reconstruction. Defining the inverse-causal rate-distortion function by

$$R_{\mathrm{IC}}(\varepsilon) = \inf I(Q; M \mid Y),$$

where M is additional information supplied to a decoder and the reconstruction error is at most ε, the high-resolution theory gives, under the regularity conditions developed above,

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{d_I(Q \mid Y)}{2} \log \frac{1}{\varepsilon} + O(1).}$$

Hence the same object appears in three forms: as the dimension of the unresolved projected fiber, as the conditional information dimension of the causal quotient, and as the leading information cost of inverse reconstruction.

The resulting hierarchy can be summarized as

$$\boxed{U \longrightarrow Q = \Phi_f(U) \longrightarrow Q \mid Y \longrightarrow R_{\mathrm{IC}}(\varepsilon).}$$

The first arrow removes latent distinctions that are causally redundant. The second records the information about the causal quotient that survives observation. The third quantifies the information that must still be supplied to reconstruct the causal antecedent.

This framework also clarifies several distinctions that are easily obscured when inverse causal inference is treated solely as a problem of conditional prediction. Causal information loss is not the same as observational uncertainty, and neither is identical to inverse information burden. A mechanism may discard distinctions within the original latent representation while retaining everything required for reversal; conversely, it may preserve the full causal latent state structurally while revealing essentially none of it through the observed outcome. The theory separates these phenomena rather than assigning them to a single notion of uncertainty.

The principal quantitative results rely on regularity conditions—smoothness, constant rank, suitable support conditions, and regular conditional laws—that make geometric dimension and information dimension coincide. The causal quotient itself is more general. This distinction is important: the quotient provides the conceptual and structural foundation of the theory, while the geometric and information-theoretic results characterize increasingly regular subclasses in which causal information loss can be measured sharply. Extending the framework to singular, fractal, rank-varying, dynamic, and unknown-mechanism settings provides a natural direction for future work.

More broadly, the paper suggests a different way of thinking about causal inference. Forward causal analysis asks what outcomes result from interventions. Inverse causal analysis asks what must be known about the mechanism and its latent state to recover the intervention or antecedent from its outcome. The answer is not determined by the ambient dimension of the latent variables, nor by observational uncertainty alone. It is determined by the causal distinctions that the mechanism preserves, the extent to which the outcome reveals those distinctions, and the information required to resolve what remains.

In this sense, inverse causal inference is fundamentally an information problem. The theory developed here provides a common language for its structural, geometric, and information-theoretic aspects:

$$\boxed{\text{causal latent complexity} = \text{causal revelation} + \text{inverse information burden.}}$$

Understanding causal mechanisms therefore requires not only knowing what they imply in the forward direction, but also understanding what information they destroy, what information they reveal, and what information remains necessary to run the causal map backward.

References

Nihat Ay and Daniel Polani. Information flows in causal networks. *Advances in Complex Systems*, 11(01):17–41, 2008.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Linear inverse problems in econometrics: Regularization methods, sequential determination of the parameter of regularization. *Handbook of Econometrics*, 6:5633–5751, 2007.

Herbert Federer. *Geometric Measure Theory*. Springer-Verlag, 1969.

Miguel A Hernán and James M Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Yingyao Hu. Identification and estimation of nonlinear models with misclassification error using instrumental variables. *Journal of Econometrics*, 144(1):282–298, 2008.

Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Dominik Janzing, David Balduzzi, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Quantifying causal influences. *The Annals of Statistics*, 41(5):2324–2358, 2013.

Tsutomu Kawabata and Amir Dembo. Rate-distortion dimension of probability measures. *IEEE Transactions on Information Theory*, 40(5):1564–1572, 1994.

John M Lee. *Introduction to Smooth Manifolds*, volume 218. Springer Science & Business Media, 2nd edition, 2012.

Charles F Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2): 319–323, 1990.

Whitney K Newey and James L Powell. Instrumental variables estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1-2):193–215, 1959.

Gabriel Schamberg and Todd P Coleman. Measuring directed information flow in continuous-time stochastic processes. *IEEE Transactions on Information Theory*, 66(1):441–461, 2020.

Lakpriya Surasinghe and Erik M Bollt. Measuring information flow via conditional correlation dimension. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(5):053133, 2020.

Yihong Wu and Sergio Verdu. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Transactions on Information Theory*, 56(8):3721–3748, 2010.

10 Appendices

11 Appendix A

12 Measurability and the Causal Quotient

This appendix establishes the measurable foundations of the causal quotient used throughout the paper. The main point is that the quotient is generated not by an arbitrary latent statistic, but by the response-function map

$$\Phi_f(u) = f_u, \qquad f_u(x) = f(x, u).$$

The quotient therefore retains precisely those distinctions in the latent state that affect the structural response over the intervention domain.

12.1 Measurability of the response-function map

Let $(\mathcal{X}, X)$, $(\mathcal{U}, U)$, and $(\mathcal{Y}, Y)$ be measurable spaces, and let

$$f : \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$$

be measurable. Let $\mathcal{Y}^{\mathcal{X}}$ denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$, equipped with the evaluation σ-field

$$E = \sigma\{e_x : x \in \mathcal{X}\},$$

where

$$e_x : \mathcal{Y}^{\mathcal{X}} \to \mathcal{Y}, \qquad e_x(h) = h(x).$$

Define

$$\Phi_f : \mathcal{U} \to \mathcal{Y}^{\mathcal{X}}, \qquad \Phi_f(u) = f_u,$$

where

$$f_u(x) = f(x, u).$$

The response-function map Φ_f is measurable from $(\mathcal{U}, U)$ to $(\mathcal{Y}^{\mathcal{X}}, E)$.

Proof. By definition, it suffices to show that the inverse image under Φ_f of every generator of E is measurable. For every $x \in \mathcal{X}$,

$$(e_x \circ \Phi_f)(u) = e_x(f_u) = f_u(x) = f(x, u).$$

Since f is measurable on $\mathcal{X} \times \mathcal{U}$, the section

$$u \mapsto f(x, u)$$

is U/Y-measurable for every fixed x. Hence

$$e_x \circ \Phi_f$$

is measurable for every $x \in \mathcal{X}$. Therefore Φ_f is measurable with respect to the evaluation σ-field E. □

12.2 The causal equivalence relation

Define a relation $\sim_f$ on $\mathcal{U}$ by

$$u \sim_f u' \iff f(x, u) = f(x, u') \quad \text{for every } x \in \mathcal{X}.$$

Thus two latent states are equivalent precisely when they induce the same structural response function over the entire intervention domain.

The relation $\sim_f$ is an equivalence relation on $\mathcal{U}$.

Proof. Reflexivity follows immediately because

$$f(x, u) = f(x, u)$$

for every $x \in \mathcal{X}$.

If $u \sim_f u'$, then

$$f(x, u) = f(x, u')$$

for every $x \in \mathcal{X}$, and equality is symmetric. Hence $u' \sim_f u$.

Finally, if $u \sim_f u'$ and $u' \sim_f u''$, then

$$f(x, u) = f(x, u') = f(x, u'')$$

for every $x \in \mathcal{X}$, so $u \sim_f u''$.

Thus $\sim_f$ is reflexive, symmetric, and transitive. □

Let

$$Q_f(U) = [U]_{\sim_f}$$

denote the resulting quotient random element. The next proposition gives the more useful representation.

For $u, u' \in \mathcal{U}$,

$$u \sim_f u' \quad \Longleftrightarrow \quad \Phi_f(u) = \Phi_f(u').$$

Consequently, the quotient $\mathcal{U}/! \sim_f$ is canonically identified with the image

$$\Phi_f(\mathcal{U}) \subseteq \mathcal{Y}^{\mathcal{X}}.$$

Proof. By definition,

$$u \sim_f u'$$

if and only if

$$f(x, u) = f(x, u') \qquad \forall x \in \mathcal{X}.$$

But

$$\Phi_f(u) = f_u \quad \text{and} \quad \Phi_f(u') = f_{u'}.$$

Hence the preceding equality holds for every x if and only if the two functions f_u and $f_{u'}$ are identical. Therefore

$$u \sim_f u' \quad \Longleftrightarrow \quad \Phi_f(u) = \Phi_f(u').$$

It follows that Φ_f is constant on equivalence classes and separates distinct equivalence classes. Define

$$\widetilde{\Phi}_f : \mathcal{U}/\sim_f \to \Phi_f(\mathcal{U})$$

by

$$\widetilde{\Phi}_f([u]) = \Phi_f(u).$$

The preceding equivalence shows that this definition is well defined. It is surjective by construction and injective because

$$\widetilde{\Phi}_f([u]) = \widetilde{\Phi}_f([u'])$$

implies

$$\Phi_f(u) = \Phi_f(u')$$

and therefore $u \sim_f u'$, so $[u] = [u']$.

Thus the quotient is canonically identified with the image of the response-function map. $\square$

The preceding result also explains why the quotient is intrinsically causal. The equivalence relation is not defined by equality of observed outcomes at a particular value of X. It requires equality of the complete response functions

$$x \mapsto f(x, u)$$

over the intervention domain. A latent distinction that changes the response under at least one admissible intervention therefore survives in the quotient.

12.3 The quotient as the minimal response-function representation

The following proposition formalizes the sense in which the quotient removes exactly the causally redundant distinctions.

Let

$$S : \mathcal{U} \to \mathcal{S}$$

be any map such that there exists a map

$$\psi : \mathcal{S} \to \mathcal{Y}^{\mathcal{X}}$$

satisfying

$$\Phi_f = \psi \circ S.$$

Then

$$S(u) = S(u') \quad \Longrightarrow \quad u \sim_f u'.$$

Equivalently, every representation through which the complete response-function map factors must distinguish every pair of latent states belonging to different causal equivalence classes.

Proof. Suppose

$$S(u) = S(u').$$

Since

$$\Phi_f = \psi \circ S,$$

we obtain

$$\Phi_f(u) = \psi(S(u)) = \psi(S(u')) = \Phi_f(u').$$

By Proposition 12.2,

$$\Phi_f(u) = \Phi_f(u') \quad \Longleftrightarrow \quad u \sim_f u'.$$

Hence the claim follows. $\square$

Thus $Q_f(U)$ is minimal in the structural sense relevant here: no representation that preserves the entire intervention-response function can merge two latent states that the quotient keeps distinct.

Notice that this is different from statistical sufficiency. The proposition does not concern a likelihood, a sampling model, or a target parameter. It concerns preservation of the structural response function itself.

12.4 Quotient-complete information

For the identification results, it is useful to distinguish the canonical quotient from an arbitrary auxiliary variable. Let

$$Q = \Phi_f(U).$$

An auxiliary random variable W is called *quotient-complete* if

$$\sigma(Q) \subseteq \sigma(W) \quad \text{modulo null sets.}$$

Under standard Borel assumptions, this is equivalent to the existence of a measurable function

$$r$$

such that

$$Q = r(W) \quad \text{a.s.}$$

Suppose Q and W take values in standard Borel spaces. Then

$$\sigma(Q) \subseteq \sigma(W) \quad \text{modulo null sets}$$

if and only if there exists a measurable map r such that

$$Q = r(W) \quad \text{a.s.}$$

Proof. This is the measurable factorization theorem for random elements on standard Borel spaces.

If $Q = r(W)$ almost surely for a measurable r, then every event in $\sigma(Q)$ is, modulo null sets, the inverse image under W of a measurable set. Hence

$$\sigma(Q) \subseteq \sigma(W)$$

modulo null sets.

Conversely, if

$$\sigma(Q) \subseteq \sigma(W)$$

modulo null sets and both random elements take values in standard Borel spaces, the Doob–Dynkin measurable factorization theorem implies the existence of a measurable function r such that

$$Q = r(W) \quad \text{a.s.}$$

□

The quotient-completeness condition is deliberately stronger than the statement that W happens to permit recovery of X under one particular probability law. An auxiliary variable may contain distribution-specific information sufficient to reconstruct X without determining the entire structural quotient. Such accidental recovery is not what the quotient is designed to characterize. Quotient completeness instead guarantees availability of the canonical latent information required to reproduce the response-function state.

12.5 Intervention-domain dependence

The causal quotient depends on the intervention domain. Let

$$\mathcal{X}_1 \subseteq \mathcal{X}_2$$

be two admissible intervention domains. Define

$$u \sim_{\mathcal{X}_j} u' \quad \Longleftrightarrow \quad f(x,u) = f(x,u') \quad \forall x \in \mathcal{X}_j.$$

Then

$$u \sim_{\mathcal{X}_2} u' \quad \Longrightarrow \quad u \sim_{\mathcal{X}_1} u',$$

because equality over the larger domain implies equality over the smaller domain.

Consequently, restricting the intervention domain can only merge causal equivalence classes; expanding the domain can only preserve or increase the distinctions regarded as causally relevant.

This observation is useful when interpreting Q as a causal latent state. The quotient is not an intrinsic property of U independently of the causal questions being asked. It is a property of the pair

$$(f, \mathcal{X}),$$

that is, of the structural mechanism together with the family of interventions under consideration.

12.6 Summary

The results of this appendix establish the foundational chain

$$U \xrightarrow{\Phi_f} Q = \Phi_f(U).$$

The map Φ_f is measurable, its fibers are exactly the structural causal equivalence classes, and the quotient is canonically represented by the response-function image. Any representation preserving the full response function must preserve these quotient distinctions. This provides the formal basis for treating Q as the canonical causally relevant latent state in the inverse problem.

13 Appendix B

14 Inverse Identification

This appendix establishes the inverse-identification results underlying the main text. The analysis separates two logically distinct sources of inverse ambiguity. The first is *structural*: even if the complete causal latent state were known, the structural map may fail to be invertible in the causal variable. The second is *informational*: the structural map may be invertible conditional on the causal quotient, but the information required to identify that quotient may not be available.

Throughout this appendix, let

$$Y = f(X,U),$$

where X takes values in a measurable space $(\mathcal{X}, X)$, U in $(\mathcal{U}, U)$, and Y in $(\mathcal{Y}, Y)$. Let

$$Q = \Phi_f(U)$$

denote the causal quotient defined in Appendix 12.

14.1 Exact inverse identification

We begin with the basic measurable characterization of exact inverse recovery.

Let X and Y be random elements. The following are equivalent:

1. There exists a measurable function $g : \mathcal{Y} \to \mathcal{X}$ such that $X = g(Y) \quad \text{a.s.}$

2. $\sigma(X) \subseteq \sigma(Y) \quad$ modulo null sets.

If $\mathcal{X}$ is standard Borel, these conditions are also equivalent to the existence of a measurable version of the conditional distribution satisfying

$$P(X \in \cdot \mid Y) = \delta_X \quad \text{a.s.}$$

Proof. If $X = g(Y)$ almost surely for a measurable g, then every event of the form

$$\{X \in A\}, \qquad A \in X,$$

is, modulo null sets, equal to

$$\{Y \in g^{-1}(A)\}.$$

Therefore

$$\sigma(X) \subseteq \sigma(Y)$$

modulo null sets.

Conversely, suppose

$$\sigma(X) \subseteq \sigma(Y)$$

modulo null sets. When $\mathcal{X}$ is standard Borel, the Doob–Dynkin measurable factorization theorem implies the existence of a measurable function g such that

$$X = g(Y) \quad \text{a.s.}$$

Finally, if $X = g(Y)$ almost surely, then conditional on Y the value of X is known almost surely, and therefore

$$P(X \in A \mid Y) = \mathbf{1}\{X \in A\} = \delta_X(A) \quad \text{a.s.}$$

for every measurable A.

Conversely, if

$$P(X \in \cdot \mid Y) = \delta_X \quad \text{a.s.},$$

then the conditional law of X given Y is degenerate almost surely. Hence X is $\sigma(Y)$-measurable, and the first condition follows. $\square$

If $X \in L^2$ and $\mathcal{X} \subseteq \mathbb{R}$, exact inverse identification from Y is equivalent to

$$\mathrm{Var}(X \mid Y) = 0 \quad \text{a.s.}$$

Proof. For square-integrable scalar X,

$$\mathrm{Var}(X \mid Y) = 0$$

almost surely if and only if

$$X = E[X \mid Y] \quad \text{a.s.}$$

Thus X is $\sigma(Y)$-measurable, and Proposition 14.1 applies. $\square$

14.2 Structural inversion conditional on the causal quotient

The next result formalizes the role of the causal quotient.

For each quotient state q in the image of Φ_f, let

$$f_q : \mathcal{X} \to \mathcal{Y}$$

denote the corresponding response function. More precisely, if

$$q = \Phi_f(u),$$

define

$$f_q(x) = f(x, u).$$

This is well defined because $\Phi_f(u) = \Phi_f(u')$ implies

$$f(x, u) = f(x, u') \qquad \forall x \in \mathcal{X}.$$

Suppose that each f_q is injective and admits a measurable inverse on its image.
Suppose that for every q in the support of Q, the response function

$$f_q : \mathcal{X} \to \mathcal{Y}$$

is injective and has a measurable inverse

$$f_q^{-1} : f_q(\mathcal{X}) \to \mathcal{X}.$$

Then

$$X = g(Y, Q) \quad \text{a.s.},$$

where

$$g(y, q) = f_q^{-1}(y).$$

In particular,

$$H(X \mid Y, Q) = 0$$

whenever the variables are discrete.

Proof. By construction,

$$Y = f_Q(X) \quad \text{a.s.}$$

where f_Q denotes the response function indexed by the realized quotient state. Since f_q is injective for every relevant q,

$$f_q^{-1}(f_q(x)) = x.$$

Consequently,

$$X = f_Q^{-1}(Y) \quad \text{a.s.}$$

Define

$$g(y, q) = f_q^{-1}(y)$$

on the set of pairs (y, q) for which $y \in f_q(\mathcal{X})$. By the assumed measurability of the family of inverses, g is measurable. Hence

$$X = g(Y, Q) \quad \text{a.s.}$$

The conditional entropy statement follows immediately in the discrete case because X is a measurable function of (Y, Q). $\square$

The theorem isolates the structural component of inverse identification. Once Q is known, there is no remaining uncertainty about X. The result does not, however, imply that Q is observed or otherwise available to the researcher. That is a separate informational question.

14.3 Quotient-complete auxiliary information

We now formalize the sufficient structural condition for an auxiliary variable to make the inverse problem solvable.

Suppose the assumptions of Theorem 14.2 hold. Let W be an auxiliary random element taking values in a standard Borel space, and suppose that W is quotient-complete:

$$\sigma(Q) \subseteq \sigma(W) \quad \text{modulo null sets.}$$

Then there exists a measurable function

$$G : \mathcal{Y} \times \mathcal{W} \to \mathcal{X}$$

such that

$$\boxed{X = G(Y, W) \quad \text{a.s.}}$$

Thus the causal antecedent is exactly identifiable from the outcome and quotient-complete auxiliary information.

Proof. By quotient completeness and the measurable factorization result of Lemma 12.4, there exists a measurable function

$$r : \mathcal{W} \to \mathcal{Q}$$

such that

$$Q = r(W) \quad \text{a.s.}$$

By Theorem 14.2, there exists a measurable function g satisfying

$$X = g(Y, Q) \quad \text{a.s.}$$

Substituting $Q = r(W)$ gives

$$X = g(Y, r(W)) \quad \text{a.s.}$$

Define

$$G(y, w) = g(y, r(w)).$$

Then G is measurable and

$$X = G(Y, W) \quad \text{a.s.}$$

as required. □

Under the assumptions of Theorem 14.3, if X, Q, W, Y are discrete, then

$$H(X \mid Y, W) = 0.$$

Moreover,

$$I(X; W \mid Y) = H(X \mid Y)$$

whenever Q itself is the additional information supplied to the decoder.

Proof. The first statement follows because X is a measurable function of (Y, W). For the second, Theorem 14.2 gives

$$H(X \mid Y, Q) = 0.$$

Therefore

$$I(X; Q \mid Y) = H(X \mid Y) - H(X \mid Y, Q) = H(X \mid Y).$$

□

14.4 What quotient completeness does and does not imply

It is important to distinguish a canonical structural requirement from distribution-specific recoverability.

Suppose the assumptions of Theorem 14.2 hold. If W is quotient-complete, then X is recoverable from (Y, W).

The converse need not hold: it is possible for

$$X = G(Y, W) \quad \text{a.s.}$$

under a particular probability law even though

$$\sigma(Q) \not\subseteq \sigma(W)$$

modulo null sets.

Proof. The first statement is Theorem 14.3.

For the second statement, it suffices to note that exact recovery of X is a statement about the realized random variable X under the given probability law, whereas quotient completeness requires preservation of the entire structural response-function state. The former can hold because the distribution of (X, Y, W) imposes additional restrictions that identify X even though W does not determine Q.

Thus exact distribution-specific recovery does not logically imply quotient completeness. □

This distinction is essential for the interpretation of the causal quotient. The quotient is intended to characterize the canonical latent information required to preserve the structural mechanism, not to provide a necessary-and-sufficient condition for every possible form of distribution-specific statistical recovery. The theory therefore uses quotient completeness as the natural structural benchmark for latent information availability.

14.5 Impossibility under structural noninvertibility

The next theorem establishes the complementary failure mode.

Suppose there exist q in the structural support of Q and distinct values

$$x_1 \neq x_2$$

such that

$$f_q(x_1) = f_q(x_2) = y.$$

Then no decoder of the form

$$X = g(Y, Q)$$

can recover X exactly on a probability law assigning positive conditional probability to both x_1 and x_2 given $Q = q$ and $Y = y$.

Proof. Suppose, toward a contradiction, that there exists a measurable function g such that

$$X = g(Y, Q) \quad \text{a.s.}$$

On the event

$$\{Q = q, Y = y\},$$

both $X = x_1$ and $X = x_2$ are possible with positive conditional probability by assumption.

But the decoder takes the same value

$$g(y, q)$$

for both realizations. It therefore cannot equal both x_1 and x_2. Hence

$$P(X \neq g(Y, Q) \mid Q = q, Y = y) > 0,$$

contradicting exact recovery. □

If the structural response function f_q is noninjective on a set of quotient states having positive probability, then no amount of additional information about Q can restore exact inversion on that set.

Proof. Knowing Q already conditions on the complete causal response function. Theorem 14.5 shows that if that response function itself maps two distinct causal antecedents to the same outcome, then X cannot be recovered from (Y, Q). Additional information that merely identifies Q therefore cannot remove the ambiguity. $\square$

14.6 The inverse-identification decomposition

We can now state the central structural decomposition.

Consider the structural causal model

$$Y = f(X, U)$$

and let

$$Q = \Phi_f(U).$$

Suppose:

1. for every relevant q, the response function f_q is injective and admits a measurable inverse;
2. W is quotient-complete, so that $\sigma(Q) \subseteq \sigma(W) \quad \text{modulo null sets.}$

Then

$$X = G(Y, W) \quad \text{a.s.}$$

for some measurable G.

Conversely, if the structural response function is noninjective on a positive-probability portion of the structural support and the conditional law of X assigns positive probability to distinct preimages of the same (Y, Q), exact structural inversion is impossible even when Q is fully known.

Thus inverse identification separates into:

$$\boxed{\text{structural invertibility} \quad + \quad \text{availability of causal quotient information.}}$$

Proof. The first statement follows directly from Theorem 14.3. The second follows from Theorem 14.5.

The two conditions concern different objects. Structural invertibility is a property of the response functions

$$f_q : \mathcal{X} \to \mathcal{Y}.$$

Quotient completeness is a property of the information available to the decoder. Hence failure of either condition constitutes a distinct source of inverse-identification failure. $\square$

14.7 Exact inverse identification from the outcome alone

The preceding results immediately characterize the special case in which no auxiliary information is available.

Suppose the assumptions of Theorem 14.2 hold. Exact recovery from Y alone,

$$X = g(Y) \quad \text{a.s.,}$$

holds if and only if

$$\sigma(X) \subseteq \sigma(Y) \quad \text{modulo null sets.}$$

Equivalently, under standard Borel assumptions,

$$P(X \in \cdot \mid Y) = \delta_X \quad \text{a.s.}$$

Proof. This is Proposition 14.1. The structural invertibility assumption ensures that the causal quotient would suffice for inversion, but it does not imply that the outcome itself contains the quotient information. Exact recovery from Y alone therefore requires the stronger condition that Y itself generate the information needed to determine X. $\square$

14.8 Discrete information requirement

For discrete variables, the preceding decomposition has a direct information-theoretic interpretation.

Suppose X, Y, Q are discrete and

$$H(X \mid Y, Q) = 0.$$

Then

$$\boxed{I(X; Q \mid Y) = H(X \mid Y).}$$

Proof. By the chain rule for conditional mutual information,

$$I(X; Q \mid Y) = H(X \mid Y) - H(X \mid Y, Q).$$

The assumed exact inverse condition gives

$$H(X \mid Y, Q) = 0.$$

Therefore

$$I(X; Q \mid Y) = H(X \mid Y).$$

$\square$

The quantity in Proposition 14.8 measures information supplied by the quotient about the antecedent after observing the outcome. It should not be confused with the information lost when the original latent variable U is compressed to Q. These are distinct quantities. The latter is governed by

$$H(U \mid Q),$$

whereas the former is governed by

$$I(X; Q \mid Y).$$

A quotient may discard substantial information about U while retaining all information relevant for inverse reconstruction.

14.9 Two logically distinct forms of information loss

The preceding results imply a useful conceptual distinction. Consider the chain

$$U \longrightarrow Q \longrightarrow X \longrightarrow Y.$$

The first transformation,

$$U \to Q,$$

removes latent distinctions that are causally redundant. The second issue concerns the information about Q that remains accessible after observing Y.

Thus there are two different questions:

1. **Causal information loss:** which distinctions in U are eliminated by the causal response-function representation? $U \to Q$.
2. **Inverse information deficit:** which causally relevant distinctions in Q remain unresolved after observing Y?

$$Q \to Q \mid Y.$$

The first is a property of the structural mechanism and intervention domain. The second depends additionally on the observational law and the way in which X and Q enter the outcome equation. This separation is the basis for the geometric and information-dimensional results in the subsequent appendices.

15 Appendix C

16 Projected-Fiber Geometry and Causal Revelation

This appendix proves the geometric results underlying the causal-revelation theory. The central object is the projected fiber of the structural map

$$Y = F(X, Q).$$

The ordinary fiber describes all pairs (x, q) compatible with an observed outcome. Because the inverse problem concerns the causal quotient Q, however, the relevant ambiguity is the projection of this fiber onto the quotient space. The dimension of this projected fiber is determined by the rank of quotient variation after removing outcome directions that can be generated by variation in X.

16.1 Setup

Let $\mathcal{X}$, $\mathcal{Q}$, and $\mathcal{Y}$ be C^r manifolds, with

$$\dim \mathcal{X} = p, \qquad \dim \mathcal{Q} = k, \qquad \dim \mathcal{Y} = m,$$

where $r \geq 1$. Let

$$F : \mathcal{X} \times \mathcal{Q} \to \mathcal{Y}$$

be a C^r map.

For a fixed outcome $y \in \mathcal{Y}$, define the fiber

$$\mathcal{F}_y = F^{-1}(y) = \{(x, q) \in \mathcal{X} \times \mathcal{Q} : F(x, q) = y\}.$$

The corresponding projected fiber is

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y) = \{q \in \mathcal{Q} : \exists x \in \mathcal{X} \text{ such that } F(x,q) = y\},$$

where

$$\pi_Q : \mathcal{X} \times \mathcal{Q} \to \mathcal{Q}$$

denotes projection onto the quotient coordinate.

The set $\mathcal{A}_y$ is the set of causal quotient states compatible with the observed outcome after allowing the causal antecedent X to vary.

16.2 The quotient derivative

Fix $(x, q) \in \mathcal{X} \times \mathcal{Q}$, and write

$$y = F(x, q).$$

The differential of F decomposes as

$$DF_{(x,q)} = \begin{bmatrix} D_x F_{(x,q)} & D_q F_{(x,q)} \end{bmatrix}.$$

The image

$$\mathcal{S}_X(x, q) = \operatorname{Im} D_x F_{(x,q)} \subseteq T_y \mathcal{Y}$$

contains the outcome directions generated by perturbations of X while holding Q fixed.

A perturbation of Q should count as revealed only if its induced outcome variation cannot be reproduced by a perturbation of X. This motivates the quotient map

$$\boxed{\overline{D}_q F_{(x,q)} : T_q \mathcal{Q} \longrightarrow T_y \mathcal{Y} / \mathcal{S}_X(x, q)}$$

defined by

$$\overline{D}_q F_{(x,q)}(v_q) = D_q F_{(x,q)} v_q + \mathcal{S}_X(x, q).$$

For $(x, q) \in \mathcal{X} \times \mathcal{Q}$,

$$\ker \overline{D}_q F_{(x,q)} = \left\{ v_q \in T_q \mathcal{Q} : \exists v_x \in T_x \mathcal{X} \text{ such that } D_x F_{(x,q)} v_x + D_q F_{(x,q)} v_q = 0 \right\}.$$

Proof. By definition,

$$v_q \in \ker \overline{D}_q F$$

if and only if

$$D_q F v_q \in \operatorname{Im} D_x F.$$

The latter condition is equivalent to the existence of some $v_x \in T_x \mathcal{X}$ satisfying

$$D_q F v_q = D_x F v_x'.$$

Taking $v_x = -v_x'$ gives

$$D_x F v_x + D_q F v_q = 0.$$

Conversely, if such a v_x exists, then

$$D_qFv_q = -D_xFv_x \in \operatorname{Im} D_xF,$$

so

$$v_q \in \ker \overline{D}_qF.$$

$\square$

Thus $\ker \overline{D}_qF$ consists exactly of quotient directions whose local effect on the outcome can be offset by an appropriate local change in X.

The *local causal revelation rank* at (x, q) is

$$\boxed{r_{\text{rev}}(x, q) = \operatorname{rank} \overline{D}_qF_{(x,q)}.}$$

The rank is not generally equal to $\operatorname{rank} D_qF$. The latter counts all local outcome directions generated by Q, including directions that are observationally confounded with changes in X. The quotient derivative removes precisely those directions.

16.3 Rank identity

The quotient derivative has a useful equivalent representation in terms of the full differential.

At every (x, q),

$$\boxed{\operatorname{rank} \overline{D}_qF = \operatorname{rank} DF - \operatorname{rank} D_xF.}$$

Proof. Let

$$A = \operatorname{Im} D_xF, \qquad B = \operatorname{Im} D_qF.$$

Then

$$\operatorname{Im} DF = A + B.$$

Hence

$$\operatorname{rank} DF = \dim(A + B).$$

The image of the quotient derivative is

$$\operatorname{Im} \overline{D}_qF = (A + B)/A.$$

Therefore

$$\operatorname{rank} \overline{D}_qF = \dim(A + B) - \dim A,$$

which gives

$$\operatorname{rank} \overline{D}_qF = \operatorname{rank} DF - \operatorname{rank} D_xF.$$

$\square$

16.4 The tangent space of the fiber

We next characterize the tangent space of a regular fiber.

Suppose y is a regular value of F. Then $\mathcal{F}_y$ is a C^r embedded submanifold of $\mathcal{X} \times \mathcal{Q}$ with

$$\dim \mathcal{F}_y = p + k - \operatorname{rank} DF.$$

Moreover, for every $(x,q) \in \mathcal{F}_y$,

$$T_{(x,q)}\mathcal{F}_y = \ker DF_{(x,q)}.$$

Proof. This is the regular level-set theorem. Since y is a regular value,

$$\operatorname{rank} DF_{(x,q)}$$

is maximal at every point of $\mathcal{F}_y$. The level set is consequently an embedded submanifold whose codimension equals the rank of DF. Hence

$$\dim \mathcal{F}_y = p + k - \operatorname{rank} DF.$$

A tangent vector

$$(v_x, v_q)$$

lies in the tangent space of the level set precisely when its first-order variation leaves F unchanged. Therefore

$$DF_{(x,q)}(v_x, v_q) = 0,$$

which gives

$$T_{(x,q)}\mathcal{F}_y = \ker DF_{(x,q)}.$$

□

16.5 Differential of the projected fiber

The key step is to determine the rank of the restriction of the projection map to the fiber.

At $(x,q) \in \mathcal{F}_y$,

$$D\pi_Q\big(T_{(x,q)}\mathcal{F}_y\big) = \ker \overline{D}_q F_{(x,q)}.$$

Consequently,

$$\boxed{\operatorname{rank}\left(D\pi_Q\big|_{T_{(x,q)}\mathcal{F}_y}\right) = k - r_{\text{rev}}(x,q).}$$

Proof. By Lemma 16.4,

$$T_{(x,q)}\mathcal{F}_y = \ker DF.$$

A vector $v_q \in T_q\mathcal{Q}$ belongs to

$$D\pi_Q(\ker DF)$$

if and only if there exists $v_x \in T_x\mathcal{X}$ such that

$$(v_x, v_q) \in \ker DF.$$

Equivalently,

$$D_x F v_x + D_q F v_q = 0.$$

By Lemma 16.2, this is equivalent to

$$v_q \in \ker \overline{D}_q F.$$

Hence

$$D\pi_Q(\ker DF) = \ker \overline{D}_q F.$$

Since

$$\dim T_q \mathcal{Q} = k,$$

rank-nullity gives

$$\dim \ker \overline{D}_q F = k - \operatorname{rank} \overline{D}_q F = k - r_{\text{rev}}.$$

Therefore the rank of the restricted projection equals

$$k - r_{\text{rev}}.$$

□

16.6 Projected-fiber dimension theorem

We can now state the main geometric result.

[Projected-Fiber Dimension Theorem] Let

$$F : \mathcal{X} \times \mathcal{Q} \to \mathcal{Y}$$

be C^r, with

$$\dim \mathcal{X} = p, \qquad \dim \mathcal{Q} = k, \qquad \dim \mathcal{Y} = m.$$

Fix an outcome $y \in \mathcal{Y}$ and a structural support

$$\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Q}.$$

Suppose:

1. F has constant rank r_F on $\mathcal{F}_y \cap \mathcal{S}$;
2. $D_x F$ has constant rank r_X on $\mathcal{F}_y \cap \mathcal{S}$;
3. the quotient derivative $\overline{D}_q F : T_q \mathcal{Q} \to T_y \mathcal{Y} / \operatorname{Im} D_x F$ has constant rank r_{rev} on $\mathcal{F}_y \cap \mathcal{S}$;
4. the restriction $\pi_Q\big|_{\mathcal{F}_y \cap \mathcal{S}}$ is a proper embedding onto its image.

Then $\mathcal{F}_y \cap \mathcal{S}$ is a properly embedded submanifold, its projected fiber

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y \cap \mathcal{S})$$

is a properly embedded submanifold of $\mathcal{Q}$, and

$$\boxed{\dim \mathcal{A}_y = k - r_{\text{rev}}.}$$

Moreover,

$$\boxed{r_{\text{rev}} = r_F - r_X.}$$

Proof. By the constant-rank level-set theorem, assumption (1) implies that

$$\mathcal{F}_y \cap \mathcal{S}$$

is an embedded submanifold of $\mathcal{X} \times \mathcal{Q}$ with dimension

$$\dim(\mathcal{F}_y \cap \mathcal{S}) = p + k - r_F.$$

At every point (x, q) of this manifold,

$$T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S}) = \ker DF_{(x,q)}.$$

Consider the restriction

$$\pi_Q\big|_{\mathcal{F}_y \cap \mathcal{S}}.$$

By Lemma 16.5,

$$D\pi_Q\left(T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S})\right) = \ker \overline{D}_q F.$$

Assumption (3) and rank-nullity therefore imply

$$\operatorname{rank}\left(D\pi_Q\big|_{T_{(x,q)}(\mathcal{F}_y \cap \mathcal{S})}\right) = k - r_{\text{rev}}.$$

By assumption (4), the restricted projection is an embedding onto its image. For an embedding, the differential has rank equal to the dimension of the image. Hence

$$\dim \mathcal{A}_y = k - r_{\text{rev}}.$$

It remains to establish the rank identity. By Lemma 16.3,

$$r_{\text{rev}} = \operatorname{rank} DF - \operatorname{rank} D_x F.$$

Under assumptions (1) and (2),

$$\operatorname{rank} DF = r_F, \qquad \operatorname{rank} D_x F = r_X,$$

so

$$r_{\text{rev}} = r_F - r_X.$$

This proves the theorem. $\square$

16.7 Alternative dimension calculation

The preceding proof can also be expressed directly through the dimensions of the fiber and the kernel of the projection. This representation makes the dimension conservation especially transparent.

Under the assumptions of Theorem 16.6,

$$\dim \mathcal{F}_y = p + k - r_F,$$

and

$$\dim \ker \left(D\pi_Q\big|_{T\mathcal{F}_y} \right) = p - r_X.$$

Consequently,

$$\dim \mathcal{A}_y = (p + k - r_F) - (p - r_X) = k - r_F + r_X.$$

Proof. The first equality follows from the constant-rank level-set theorem.

For the second, the kernel of the restricted projection consists of tangent vectors of the form

$$(v_x, 0)$$

that satisfy

$$D_x F v_x = 0.$$

Thus

$$\ker \left(D\pi_Q\big|_{T\mathcal{F}_y} \right) \cong \ker D_x F.$$

By rank-nullity,

$$\dim \ker D_x F = p - r_X.$$

Therefore rank-nullity for the restricted projection gives

$$\dim \mathcal{A}_y = \dim \mathcal{F}_y - (p - r_X),$$

which equals

$$(p + k - r_F) - (p - r_X) = k - r_F + r_X.$$

Using

$$r_{\text{rev}} = r_F - r_X$$

then gives

$$\dim \mathcal{A}_y = k - r_{\text{rev}}.$$

□

16.8 Interpretation of revelation rank

The projected-fiber theorem yields an immediate interpretation of the revelation rank.

Under the assumptions of Theorem 16.6:

1. If $r_{\text{rev}} = 0$, then $\dim \mathcal{A}_y = k$. Locally, the outcome does not reduce the dimension of the compatible causal quotient states.

   “‘

2. If

   $$r_{\text{rev}} = k,$$

   then

   $$\dim \mathcal{A}_y = 0.$$

   Locally, the outcome eliminates all continuous quotient ambiguity.

3. If

   $$0 < r_{\text{rev}} < k,$$

   then exactly r_{rev} local quotient dimensions are revealed and

   $$k - r_{\text{rev}}$$

   remain unresolved. “‘

Proof. Each statement follows directly from

$$\dim \mathcal{A}_y = k - r_{\text{rev}}.$$

□

This gives the local dimension-conservation identity

$$\boxed{k = r_{\text{rev}} + \dim \mathcal{A}_y.}$$

The left-hand side is the local dimension of the causal quotient, the first term is the dimension revealed by the observation, and the second is the dimension remaining compatible with the observation.

16.9 Canonical examples

16.9.1 Additive confounding

Consider

$$Y = X + Q, \qquad X, Q \in \mathbb{R}.$$

Then

$$F(x, q) = x + q,$$

and

$$D_x F = 1, \qquad D_q F = 1.$$

Hence

$$\operatorname{Im} D_x F = \mathbb{R},$$

so the quotient space

$$T_y \mathcal{Y} / \operatorname{Im} D_x F$$

is zero-dimensional. Consequently,

$$\overline{D}_q F = 0, \qquad r_{\text{rev}} = 0.$$

The projected fiber is

$$\mathcal{A}_y = \{q : y = x + q \text{ for some } x\} = \mathbb{R},$$

so

$$\dim \mathcal{A}_y = 1.$$

Thus observing Y reveals no local quotient dimension.

16.9.2 Direct observation of the quotient

Consider

$$Y = (X, Q).$$

Then

$$F(x, q) = (x, q).$$

The image of $D_x F$ consists only of directions in the X coordinate, while $D_q F$ spans the quotient coordinate. Hence

$$r_{\text{rev}} = k,$$

and

$$\dim \mathcal{A}_y = 0.$$

The quotient is locally fully revealed.

16.9.3 Partial revelation

Let

$$X = (X_1, X_2), \qquad Q = (Q_1, Q_2),$$

and consider

$$Y = (X_1 + Q_1, Q_2).$$

Then

$$D_x F = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad D_q F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The Q_1 direction lies in the image generated by X_1, whereas the Q_2 direction is directly revealed. Therefore

$$r_{\text{rev}} = 1.$$

Since $k = 2$,

$$\dim \mathcal{A}_y = 1.$$

The observation reveals one quotient dimension and leaves one unresolved.

16.10 Local revelation versus global identification

The revelation rank is intrinsically local. It characterizes the tangent dimension of the projected fiber under the regularity assumptions of the theorem. It therefore does not by itself establish global uniqueness.

Suppose that

$$r_{\text{rev}} = k$$

on a regular projected fiber. Then $\mathcal{A}_y$ is locally zero-dimensional. It may nevertheless contain more than one isolated point.

Proof. By Theorem 16.6,

$$\dim \mathcal{A}_y = 0.$$

A zero-dimensional embedded manifold is locally a discrete collection of points. Nothing in the local rank condition requires the global projected fiber to contain only one such point. Hence multiple isolated compatible quotient states may remain. □

This distinction is particularly important for nonlinear mechanisms with discrete symmetries. For example, a transformation may reveal all continuous quotient dimensions while leaving a finite ambiguity such as q versus $-q$. Information dimension treats such a finite ambiguity as dimension zero, even though exact global identification still requires discrete information.

16.11 Why the projected fiber is the relevant object

The full fiber

$$\mathcal{F}_y = \{(x, q) : F(x, q) = y\}$$

contains two sources of ambiguity: variation in the causal antecedent X and variation in the causal quotient Q. Its dimension therefore does not directly measure the inverse uncertainty about Q.

The projection

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y)$$

removes the distinction between different values of X that are compatible with the same quotient state. Consequently,

$$\mathcal{A}_y$$

is the geometric identified set for the causal quotient.

The quotient derivative provides the corresponding infinitesimal construction. A quotient perturbation v_q is unresolved precisely when its effect on the outcome can be offset by some perturbation v_x. In symbols,

$$v_q \in \ker \overline{D}_q F \quad \Longleftrightarrow \quad \exists v_x : D_x F v_x + D_q F v_q = 0.$$

Thus

$$\ker \overline{D}_q F$$

is the tangent space of locally unresolved quotient directions.

16.12 Relation to causal versus observational information

The geometric construction begins with the causal quotient

$$Q = \Phi_f(U).$$

This determines which latent distinctions are causally meaningful. The projected fiber then asks which of those distinctions remain compatible with an observed outcome.

Accordingly, the two stages answer different questions:

$$\boxed{\Phi_f : \text{Which latent distinctions matter causally?}}$$

and

$$\boxed{P(Q \mid Y) : \text{Which of those distinctions are revealed observationally?}}$$

The differential object

$$\overline{D}_q F$$

is the local geometric bridge between these two questions. Its rank measures how many causal-quotient directions the observation locally separates after accounting for the freedom to vary X.

16.13 Summary

The results of this appendix establish the geometric core of the theory. Under constant-rank and embedding conditions, the projected fiber of the causal quotient has dimension

$$\boxed{\dim \mathcal{A}_y = k - r_{\mathrm{rev}},}$$

where

$$\boxed{r_{\mathrm{rev}} = \operatorname{rank} \overline{D}_q F = \operatorname{rank} DF - \operatorname{rank} D_x F.}$$

Thus the local causal quotient dimension decomposes as

$$\boxed{k = \underbrace{r_{\mathrm{rev}}}_{\text{revealed}} + \underbrace{\dim \mathcal{A}_y}_{\text{unresolved}}.}$$

This geometric identity provides the foundation for the conditional information-dimension result proved in Appendix 18.

17 Appendix D

18 Conditional Information Dimension and Causal Revelation

This appendix connects the geometric projected-fiber result of Appendix 16 to conditional information dimension. The central result shows that, under suitable regularity conditions on the conditional law, the dimension of the unresolved causal quotient is equal to the conditional information dimension of the quotient given the observed outcome.

The argument proceeds in three steps. First, the conditional distribution of (X, Q) given $Y = y$ is supported on the structural fiber $F^{-1}(y)$. Second, its Q-marginal is therefore supported on the projected fiber $\mathcal{A}_y$. Third, when this projected conditional law is sufficiently regular with respect to Hausdorff measure on $\mathcal{A}_y$, its information dimension equals the manifold dimension of $\mathcal{A}_y$.

18.1 Conditional laws on structural fibers

Let

$$(X, Q) \sim \mu$$

on $\mathcal{X} \times \mathcal{Q}$, and let

$$Y = F(X, Q),$$

where F is measurable. Let ν denote the distribution of Y.

Assume that a regular conditional distribution

$$\mu_y = P_{X,Q|Y=y}$$

exists for ν-almost every y. Define the projected conditional law

$$\lambda_y = P_{Q|Y=y} = (\pi_Q)_{\#}\mu_y.$$

For ν-almost every y,

$$\mu_y\big(F^{-1}(y)\big) = 1.$$

Consequently,

$$\lambda_y(\mathcal{A}_y) = 1,$$

where

$$\mathcal{A}_y = \pi_Q(F^{-1}(y)).$$

Proof. Since

$$Y = F(X, Q) \quad \text{a.s.},$$

we have

$$P\big(F(X, Q) = Y\big) = 1.$$

Equivalently,

$$P\big((X,Q) \in F^{-1}(Y)\big) = 1.$$

By the defining property of a regular conditional distribution,

$$\mu_y\big(F^{-1}(y)\big) = 1$$

for ν-almost every y.
Now let

$$\mathcal{A}_y = \pi_Q(F^{-1}(y)).$$

Since

$$F^{-1}(y) \subseteq \pi_Q^{-1}(\mathcal{A}_y),$$

we obtain

$$\mu_y\big(\pi_Q^{-1}(\mathcal{A}_y)\big) = 1.$$

By definition of pushforward measure,

$$\lambda_y(\mathcal{A}_y) = \mu_y\big(\pi_Q^{-1}(\mathcal{A}_y)\big) = 1.$$

□

The result is purely structural. It does not require a density for (X, Q) or for the conditional law. The conditional distribution is supported on the structural fiber simply because the observed outcome is generated by the structural equation.

18.2 Information dimension on a regular manifold

We next state the information-dimension result needed for the main theorem.

For a random vector Z taking values in $\mathbb{R}^d$, define the uniform scalar quantization

$$[Z]_\delta = \delta \left\lfloor \frac{Z}{\delta} \right\rfloor.$$

Its lower and upper information dimensions are

$$\underline{d}_I(Z) = \liminf_{\delta \downarrow 0} \frac{H([Z]_\delta)}{\log(1/\delta)},$$

and

$$\overline{d}_I(Z) = \limsup_{\delta \downarrow 0} \frac{H([Z]_\delta)}{\log(1/\delta)},$$

whenever the entropies are well defined.

For conditional random variables, define

$$\underline{d}_I(Q \mid Y = y) = \liminf_{\delta \downarrow 0} \frac{H([Q]_\delta \mid Y = y)}{\log(1/\delta)}$$

and analogously for the upper information dimension.

When the conditional distribution is supported on a smooth manifold, it is convenient to work in local coordinates.

[Information Dimension of a Regular Manifold Law] Let $\mathcal{M} \subseteq \mathbb{R}^m$ be a d-dimensional C^1 embedded manifold, and let Z be supported on $\mathcal{M}$. Suppose that, locally on $\mathcal{M}$, the law of Z is absolutely continuous with respect to

$$\mathcal{H}^d\big|_{\mathcal{M}}$$

with a density that is positive and finite almost everywhere and satisfies the local regularity conditions required for information dimension to be invariant under bi-Lipschitz coordinate maps. Then

$$\boxed{d_I(Z) = d.}$$

Proof. Fix a point of $\mathcal{M}$. By the embedded-manifold property, there exists a neighborhood $V \subseteq \mathcal{M}$ and a C^1 coordinate chart

$$\psi : U \subseteq \mathbb{R}^d \to V.$$

After restricting to a sufficiently small neighborhood, ψ and ψ^{-1} are bi-Lipschitz.
Let

$$T = \psi^{-1}(Z)$$

on the event $Z \in V$. Because the law of Z is absolutely continuous with respect to $\mathcal{H}^d|_{\mathcal{M}}$ with positive finite density, the coordinate representation T has a law absolutely continuous with respect to d-dimensional Lebesgue measure, with a locally positive and finite density after accounting for the chart Jacobian.

A d-dimensional random vector with an absolutely continuous law satisfying the stated regularity conditions has information dimension d. Hence

$$d_I(T) = d.$$

Information dimension is invariant under locally bi-Lipschitz transformations. Since

$$Z = \psi(T),$$

it follows that

$$d_I(Z) = d$$

locally. A finite or countable regular chart cover then gives the same dimension globally under the assumed regularity conditions. □

The regularity assumption in Lemma 18.2 is substantive. Absolute continuity with respect to Hausdorff measure alone does not automatically provide every integrability or entropy condition needed for a particular quantization definition. The theorem is therefore stated for the regularity class in which information dimension is known to be invariant under the relevant coordinate maps.

18.3 Conditional information dimension of the projected fiber

We can now combine the support result with the geometric theorem.

[Conditional Information-Dimension Revelation Theorem] Suppose the assumptions of Theorem 16.6 hold for ν-almost every y, and let

$$d = k - r_{\text{rev}}.$$

Suppose further that, for ν-almost every y:

1. $\mathcal{A}_y$ is a d-dimensional C^1 embedded manifold;
2. the conditional quotient law $\lambda_y = P_{Q|Y=y}$ satisfies $\lambda_y \ll \mathcal{H}^d\big|_{\mathcal{A}_y}$;
3. the Radon–Nikodym density is positive and finite $\mathcal{H}^d$-almost everywhere and satisfies the local regularity conditions of Lemma 18.2.

Then

$$\boxed{d_I(Q \mid Y = y) = k - r_{\text{rev}}}$$

for ν-almost every y.
If, in addition,

$$d_I(Q) = k,$$

then

$$\boxed{d_I(Q) - d_I(Q \mid Y) = r_{\text{rev}}}$$

whenever the conditional information dimension is well defined almost surely and admits the stated aggregation.

Proof. By Lemma 18.1,

$$\lambda_y(\mathcal{A}_y) = 1$$

for ν-almost every y.
By Theorem 16.6,

$$\dim \mathcal{A}_y = k - r_{\text{rev}} = d.$$

Thus λ_y is supported on a d-dimensional embedded manifold.

By assumptions (2) and (3), the conditional law is regular with respect to the d-dimensional Hausdorff measure on that manifold. Lemma 18.2 therefore gives

$$d_I(Q \mid Y = y) = d.$$

Substituting

$$d = k - r_{\text{rev}}$$

yields

$$d_I(Q \mid Y = y) = k - r_{\text{rev}}.$$

If

$$d_I(Q) = k,$$

then

$$d_I(Q) - d_I(Q \mid Y) = k - (k - r_{\text{rev}}) = r_{\text{rev}},$$

provided the conditional dimension aggregates to the unconditional conditional information dimension in the stated regularity class. □

18.4 A chart-based proof

For completeness, the preceding theorem can be expressed directly in local coordinates. This makes explicit why the dimension is intrinsic to the projected fiber rather than to its ambient representation.

Under the assumptions of Theorem 18.3, for ν-almost every y there exists a local chart

$$\psi_y : V_y \subseteq \mathbb{R}^d \to \mathcal{A}_y$$

such that, locally,

$$Q = \psi_y(T_y),$$

where the conditional law of T_y given $Y = y$ is absolutely continuous with respect to d-dimensional Lebesgue measure. Consequently,

$$d_I(Q \mid Y = y) = d.$$

Proof. Since $\mathcal{A}_y$ is a d-dimensional embedded manifold, every point of $\mathcal{A}_y$ has a neighborhood admitting a coordinate chart

$$\psi_y : V_y \subseteq \mathbb{R}^d \to \mathcal{A}_y.$$

After restricting the chart neighborhood, both ψ_y and its inverse are bi-Lipschitz.

The conditional law of Q is absolutely continuous with respect to

$$\mathcal{H}^d|_{\mathcal{A}_y}.$$

The area formula for a C^1 chart implies that the pullback law under

$$T_y = \psi_y^{-1}(Q)$$

is absolutely continuous with respect to Lebesgue measure on V_y. Under the assumed positive, finite, and regular density conditions,

$$d_I(T_y) = d.$$

Bi-Lipschitz invariance of information dimension then gives

$$d_I(Q \mid Y = y) = d_I(T_y) = d.$$

□

18.5 Coarea representation

The preceding result can be connected explicitly to the conditional density generated by a smooth structural mechanism.

Suppose now that

$$(X, Q) \in \mathbb{R}^{p+k}$$

has a joint density

$$p_{X,Q}(x, q),$$

and that

$$F : \mathbb{R}^{p+k} \to \mathbb{R}^m$$

is sufficiently smooth and has constant rank r_F on the relevant support.

Let

$$J_{r_F} F(x,q)$$

denote the r_F-dimensional Jacobian of F.

Under the constant-rank and regularity conditions required by the coarea formula, the conditional law of (X,Q) given $Y=y$ has density with respect to Hausdorff measure on the level set

$$\mathcal{F}_y = F^{-1}(y)$$

proportional to

$$\boxed{\frac{p_{X,Q}(x,q)}{J_{r_F} F(x,q)}}.$$

More precisely, for integrable test functions φ,

$$E[\varphi(X,Q) \mid Y=y] = \frac{\displaystyle\int_{\mathcal{F}_y} \varphi(x,q) \frac{p_{X,Q}(x,q)}{J_{r_F} F(x,q)} \, d\mathcal{H}^{p+k-r_F}(x,q)}{\displaystyle\int_{\mathcal{F}_y} \frac{p_{X,Q}(x,q)}{J_{r_F} F(x,q)} \, d\mathcal{H}^{p+k-r_F}(x,q)}$$

for almost every y for which the denominator is positive and finite.

Proof. The coarea formula for a constant-rank map gives

$$\int_{\mathbb{R}^{p+k}} \varphi(x,q) p_{X,Q}(x,q) \, d(x,q)$$

as an integral over the level sets of F with respect to the Hausdorff measure

$$d\mathcal{H}^{p+k-r_F}.$$

Specifically, the disintegration of the joint law along the fibers is weighted by the inverse Jacobian factor

$$\frac{1}{J_{r_F} F(x,q)}.$$

Normalizing the resulting fiber measure by its total mass yields the regular conditional distribution given $Y=y$. This gives the stated expression. □

The coarea representation is not needed for the dimension identity itself. Its role is to provide a convenient sufficient condition for the conditional quotient law to inherit regularity from a smooth joint density. In particular, if the resulting projected law has a positive finite density with respect to $\mathcal{H}^d|_{\mathcal{A}_y}$, the assumptions of Theorem 18.3 are satisfied.

18.6 Revelation as a dimension reduction

The conditional information-dimension theorem yields a natural definition of causal revelation.

Suppose $d_I(Q)$ and $d_I(Q \mid Y)$ exist. The *inverse-causal revelation* of Y about the causal quotient is

$$\boxed{\mathcal{R}_{\mathrm{IC}} = d_I(Q) - d_I(Q \mid Y).}$$

Under the assumptions of Theorem 18.3, if

$$d_I(Q) = k,$$

then

$$\boxed{\mathcal{R}_{\mathrm{IC}} = r_{\mathrm{rev}}.}$$

Thus the differential rank has an information-theoretic interpretation: it measures the number of quotient dimensions locally revealed by the observation.

18.7 Examples

18.7.1 No quotient revelation

Consider

$$Y = X + Q,$$

where $X, Q \in \mathbb{R}$ and Q has a regular one-dimensional conditional law given Y. From Appendix 16,

$$r_{\mathrm{rev}} = 0.$$

Hence

$$d_I(Q \mid Y) = 1$$

under the regularity assumptions, and therefore

$$\mathcal{R}_{\mathrm{IC}} = 0.$$

The outcome constrains X and Q jointly but does not reduce the local dimension of the compatible quotient state.

18.7.2 Complete quotient revelation

Consider

$$Y = (X, Q),$$

with $Q \in \mathbb{R}^k$. Then

$$r_{\mathrm{rev}} = k,$$

so

$$d_I(Q \mid Y) = 0.$$

Hence

$$\mathcal{R}_{\mathrm{IC}} = k$$

when $d_I(Q) = k$.

The outcome reveals the entire continuous quotient state.

18.7.3 Partial revelation

Consider

$$Y = (X_1 + Q_1, Q_2),$$

with

$$X, Q \in \mathbb{R}^2.$$

As shown in Appendix 16,

$$r_{\mathrm{rev}} = 1.$$

Under regularity,

$$d_I(Q \mid Y) = 1.$$

Thus one of the two quotient dimensions is revealed and one remains unresolved:

$$2 = 1 + 1.$$

18.8 Dimension does not capture finite ambiguity

The information-dimension result concerns continuous degrees of freedom. It therefore does not detect finite multiplicity.

Suppose that for almost every y, the projected fiber $\mathcal{A}_y$ consists of finitely many points. Then

$$d_I(Q \mid Y = y) = 0$$

under the standard quantization definition, even when

$$\#\mathcal{A}_y > 1.$$

Proof. A finite-valued random variable has bounded quantization entropy:

$$H([Q]_\delta \mid Y = y) \leq \log \#\mathcal{A}_y$$

for sufficiently small δ. Therefore

$$0 \leq \frac{H([Q]_\delta \mid Y = y)}{\log(1/\delta)} \leq \frac{\log \#\mathcal{A}_y}{\log(1/\delta)} \longrightarrow 0.$$

Hence

$$d_I(Q \mid Y = y) = 0.$$

□

Proposition 22.8 shows why information dimension should not be interpreted as a complete measure of global identification complexity. A zero-dimensional projected fiber may still contain multiple isolated structural possibilities. Such discrete ambiguity requires a separate notion of identification complexity, whereas information dimension measures the continuous component of inverse uncertainty.

18.9 Main geometric-information identity

Combining the results of Appendices 16 and 18 gives the central bridge of the paper.

[Geometric-Information Revelation Identity] Suppose that the assumptions of Theorem 16.6 and Theorem 18.3 hold, and suppose

$$d_I(Q) = k.$$

Then

$$\boxed{d_I(Q \mid Y) = k - r_{\mathrm{rev}},}$$

and

$$\boxed{d_I(Q) - d_I(Q \mid Y) = r_{\mathrm{rev}}.}$$

Equivalently,

$$\boxed{\text{causal quotient dimension} = \text{revealed quotient dimension} + \text{unresolved quotient dimension.}}$$

Proof. The first equality is Theorem 18.3:

$$d_I(Q \mid Y) = k - r_{\mathrm{rev}}.$$

Since

$$d_I(Q) = k,$$

subtraction gives

$$d_I(Q) - d_I(Q \mid Y) = k - (k - r_{\mathrm{rev}}) = r_{\mathrm{rev}}.$$

The final identity is the same equality expressed in terms of the three components. □

18.10 Causal meaning of the information-dimension identity

The identity in Theorem 18.9 should be interpreted in two stages. The causal quotient

$$Q = \Phi_f(U)$$

is defined structurally, before conditioning on the observed outcome. It identifies the latent distinctions that matter for the response of the mechanism over the intervention domain. The conditional law

$$P(Q \mid Y)$$

then measures which of those causally relevant distinctions remain unresolved after observation. Accordingly,

$$\Phi_f$$

and

$$P(Q \mid Y)$$

play fundamentally different roles:

$$\boxed{\Phi_f \ : \ \text{causal relevance,}}$$

$$\boxed{P(Q \mid Y) \ : \ \text{observational revelation.}}$$

The geometric revelation rank connects these two levels locally. It measures how many quotient directions the observation separates after accounting for outcome variation that can already be generated by the causal variable X.

This establishes the bridge needed for the rate-distortion results in Appendix 20: the dimension of the unresolved projected fiber is not merely a geometric descriptor, but the asymptotic dimension of the information that must be supplied to reverse the causal mechanism.

19 Appendix E

20 Inverse-Causal Rate Distortion and High-Resolution Reconstruction

This appendix develops the information-theoretic reconstruction results underlying the inverse-causal rate-distortion framework. The central object is the conditional information required to reconstruct the causal antecedent X once the observed outcome Y is available. Because Y is treated as freely available side information, the relevant rate measures only additional information about the causal quotient that must be supplied to the decoder.

The analysis proceeds in three stages. First, we define the inverse-causal rate-distortion function. Second, we establish its connection to conditional rate-distortion for the causal quotient. Third, under smooth invertibility and regularity conditions, we derive the high-resolution asymptotic rate and its information-dimension interpretation.

20.1 Inverse-causal rate distortion

Let

$$Y = F(X, Q),$$

and suppose that X can be reconstructed from (Y, Q) through a measurable decoder

$$X = g(Y, Q).$$

Let

$$d_X : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \infty)$$

be a distortion function.

An auxiliary representation M is generated from (Q, Y) according to a conditional distribution

$$P_{M|Q,Y}.$$

A decoder observes (Y, M) and produces

$$\widehat{X} = \widehat{g}(Y, M).$$

The *inverse-causal rate-distortion function* is

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \inf I(Q; M \mid Y),}$$

where the infimum is taken over all conditional representations $P_{M|Q,Y}$ and measurable decoders $\widehat{g}$ satisfying

$$E[d_X(X, \widehat{X})] \le \varepsilon.$$

The conditioning on Y is essential. The outcome is assumed to be observed without information cost. The rate therefore measures only the additional information about the causally relevant latent state that is required to reverse the mechanism to distortion level ε.

20.2 Exact reconstruction

The exact-reconstruction limit provides the discrete counterpart of the inverse information requirement.

Suppose X, Q, Y are discrete and

$$H(X \mid Y, Q) = 0.$$

Then the minimum additional information required for exact reconstruction is

$$\boxed{\inf I(Q; M \mid Y) = H(X \mid Y),}$$

where the infimum is over representations M from which X can be recovered exactly together with Y.

Proof. Because X is a measurable function of (Y, Q),

$$H(X \mid Y, Q) = 0.$$

Taking

$$M = Q$$

gives exact reconstruction and therefore

$$I(Q; M \mid Y) = I(Q; Q \mid Y) = H(Q \mid Y).$$

This quantity need not equal $H(X \mid Y)$ if Q contains redundant information relative to the realized X.

To characterize the minimum over arbitrary representations, note that exact reconstruction implies

$$H(X \mid Y, M) = 0.$$

Hence

$$I(X; M \mid Y) = H(X \mid Y).$$

Since X is a measurable function of (Y, Q) and M is generated from (Q, Y),

$$X - (Q, Y) - M$$

forms a conditional Markov chain. The data-processing inequality gives

$$I(X; M \mid Y) \leq I(Q; M \mid Y).$$

Therefore

$$I(Q; M \mid Y) \geq H(X \mid Y).$$

The lower bound is attainable whenever a representation of Q can preserve exactly the information about X relevant conditional on Y. Thus the minimum additional information required for exact reconstruction is

$$H(X \mid Y).$$

□

The proposition concerns the minimum information required to reconstruct X, not necessarily the entropy of the entire causal quotient. In the structural framework, Q is the canonical causally relevant latent state, while a rate-optimal representation may compress Q further when only reconstruction of the realized X is required. The inverse-causal rate-distortion formulation therefore measures task-specific reconstruction complexity conditional on the observed outcome.

20.3 Quotient reconstruction

For the high-resolution theory it is useful to distinguish reconstruction of Q from reconstruction of X.

Suppose first that

$$Q \in \mathbb{R}^k$$

and consider squared-error distortion

$$d_Q(q, \widehat{q}) = \|q - \widehat{q}\|^2.$$

Define the conditional rate-distortion function

$$R_{Q|Y}(D) = \inf I(Q; \widehat{Q} \mid Y),$$

where the infimum is over conditional reconstruction kernels satisfying

$$E\|Q - \widehat{Q}\|^2 \leq D.$$

Suppose that

$$X = g(Y, Q).$$

If, locally, $g(y, \cdot)$ is bi-Lipschitz, then small quotient reconstruction error is equivalent, up to constant factors, to small inverse reconstruction error.

Suppose that on the relevant support there exist constants

$$0 < c \leq C < \infty$$

such that

$$c\|q-q'\| \le \|g(y,q)-g(y,q')\| \le C\|q-q'\|$$

for all relevant (y,q,q'). Then

$$c^2\|Q-\widehat{Q}\|^2 \le \|X-g(Y,\widehat{Q})\|^2 \le C^2\|Q-\widehat{Q}\|^2.$$

Consequently, the corresponding rate-distortion functions have the same high-resolution logarithmic coefficient.

Proof. The two inequalities follow immediately by applying the assumed bi-Lipschitz bounds to the realized values of $(Y,Q,\widehat{Q})$ and squaring.

If

$$E\|Q-\widehat{Q}\|^2 \le D,$$

then

$$E\|X-g(Y,\widehat{Q})\|^2 \le C^2 D.$$

Conversely, if

$$E\|X-g(Y,\widehat{Q})\|^2 \le \varepsilon,$$

then

$$E\|Q-\widehat{Q}\|^2 \le \frac{\varepsilon}{c^2}.$$

Replacing the distortion argument by a fixed multiplicative constant changes a high-resolution rate only by an $O(1)$ term because

$$\log\frac{1}{C^2\varepsilon} = \log\frac{1}{\varepsilon} + O(1).$$

Thus the leading logarithmic coefficient is identical. □

20.4 Conditional information dimension

For a random vector $Q \in \mathbb{R}^k$, define

$$[Q]_\delta = \delta\left\lfloor \frac{Q}{\delta} \right\rfloor.$$

The lower and upper conditional information dimensions are

$$\underline{d}_I(Q \mid Y) = \liminf_{\delta\downarrow 0} \frac{H([Q]_\delta \mid Y)}{\log(1/\delta)},$$

and

$$\overline{d}_I(Q \mid Y) = \limsup_{\delta\downarrow 0} \frac{H([Q]_\delta \mid Y)}{\log(1/\delta)}.$$

When these agree, denote their common value by

$$d_I(Q \mid Y).$$

The following theorem states the rate-dimension bridge used in the paper.

[Conditional Rate-Dimension Theorem] Suppose that $Q \in \mathbb{R}^k$ has finite second moment and that the conditional information dimension

$$d_I(Q \mid Y)$$

exists. Suppose further that the conditional law belongs to a regularity class for which high-resolution conditional rate-distortion is asymptotically equivalent to conditional quantization entropy. Then, under squared-error distortion,

$$\boxed{\lim_{D \downarrow 0} \frac{2R_{Q|Y}(D)}{\log(1/D)} = d_I(Q \mid Y).}$$

More generally,

$$\liminf_{D \downarrow 0} \frac{2R_{Q|Y}(D)}{\log(1/D)} \geq \underline{d}_I(Q \mid Y),$$

and

$$\limsup_{D \downarrow 0} \frac{2R_{Q|Y}(D)}{\log(1/D)} \leq \overline{d}_I(Q \mid Y)$$

whenever the corresponding regularity bounds hold.

Proof. We give the standard high-resolution argument in conditional form.

For the upper bound, let

$$\widehat{Q} = [Q]_\delta$$

with

$$\delta = \sqrt{D/c_0}$$

for a fixed constant c_0 depending only on the chosen scalar quantizer and dimension. Then

$$E\|Q - \widehat{Q}\|^2 \leq D.$$

Consequently,

$$R_{Q|Y}(D) \leq I(Q; \widehat{Q} \mid Y) \leq H(\widehat{Q} \mid Y).$$

Therefore

$$R_{Q|Y}(D) \leq H([Q]_\delta \mid Y).$$

Since

$$\log(1/\delta) = \frac{1}{2}\log(1/D) + O(1),$$

we obtain

$$\limsup_{D \downarrow 0} \frac{2R_{Q|Y}(D)}{\log(1/D)} \leq \overline{d}_I(Q \mid Y).$$

For the lower bound, choose

$$\delta = D^{1/2-\eta}$$

for any fixed

$$0 < \eta < 1/2.$$

If

$$E\|Q - \widehat{Q}\|^2 \le D,$$

Markov's inequality gives

$$P(\|Q - \widehat{Q}\| > \delta) \le \frac{D}{\delta^2} = D^{2\eta}.$$

Thus, with probability tending to one, the reconstruction lies in the same δ-scale quantization neighborhood as Q. Standard entropy decomposition and Fano-type arguments then imply

$$I(Q; \widehat{Q} \mid Y) \ge H([Q]_\delta \mid Y) - o(\log(1/D)).$$

Taking the infimum over all admissible reconstructions gives

$$R_{Q|Y}(D) \ge H([Q]_\delta \mid Y) - o(\log(1/D)).$$

Because

$$\log(1/\delta) = \left(\frac{1}{2} - \eta\right) \log(1/D),$$

we obtain a lower bound approaching the conditional information dimension as $\eta \downarrow 0$. Under the assumed regularity conditions, the resulting liminf is

$$\underline{d}_I(Q \mid Y).$$

When the lower and upper dimensions coincide, the two bounds yield

$$\lim_{D \downarrow 0} \frac{2R_{Q|Y}(D)}{\log(1/D)} = d_I(Q \mid Y).$$

□

The theorem is intentionally stated with a regularity qualification. Information dimension alone does not guarantee every technical property needed to identify the exact rate-distortion asymptotics for arbitrary distributions. The dimension gives the high-resolution coefficient within the standard regularity class in which quantization entropy and rate-distortion are asymptotically equivalent.

20.5 Inverse-causal rate dimension

We now return to the inverse problem.

The *inverse-causal information dimension* is

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = 2 \lim_{\varepsilon \downarrow 0} \frac{R_{\mathrm{IC}}(\varepsilon)}{\log(1/\varepsilon)},}$$

whenever the limit exists.

[Inverse-Causal Rate-Dimension Theorem] Suppose that

$$X = g(Y, Q),$$

where Q takes values in a regular k-dimensional representation space, and suppose that for almost every y, the map

$$q \mapsto g(y, q)$$

is locally bi-Lipschitz on the relevant conditional support. Assume that the conditional rate-dimension theorem applies to $Q \mid Y$. Then

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = d_I(Q \mid Y).}$$

In particular, if

$$d_I(Q \mid Y) = d,$$

then

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{d}{2} \log \frac{1}{\varepsilon} + o\left(\log \frac{1}{\varepsilon}\right).}$$

Proof. By Lemma 20.3, the inverse reconstruction distortion induced by

$$\widehat{X} = g(Y, \widehat{Q})$$

is locally equivalent, up to multiplicative constants, to quotient distortion

$$\|Q - \widehat{Q}\|^2.$$

Consequently, the inverse-causal rate-distortion function and the conditional quotient rate-distortion function differ only by a change in the distortion argument by bounded multiplicative factors and by terms that do not affect the leading logarithmic coefficient.

Hence

$$R_{\mathrm{IC}}(\varepsilon) = R_{Q|Y}(c\varepsilon) + O(1)$$

at the level of high-resolution logarithmic asymptotics, for suitable positive constants c.

Applying Theorem 20.4,

$$\lim_{\varepsilon \downarrow 0} \frac{2R_{\mathrm{IC}}(\varepsilon)}{\log(1/\varepsilon)} = d_I(Q \mid Y).$$

Therefore

$$\mathfrak{d}_{\mathrm{IC}} = d_I(Q \mid Y).$$

The second statement follows immediately. □

20.6 High-resolution smooth reconstruction

The preceding theorem is formulated in terms of information dimension. Under stronger smoothness conditions, the coefficient can be identified directly with the dimension of the quotient.

[High-Resolution Inverse Reconstruction] Suppose:

1. $Q \in \mathbb{R}^k$;
2. $X = g(Y, Q)$;

3. for almost every y, $g(y, \cdot)$ is C^1;

4. the Jacobian $J_q g(y, q)$ has full column rank k on the relevant support;

5. the conditional law of $Q \mid Y$ is regular with a positive finite density;

6. the required moment, integrability, and high-resolution regularity conditions hold;

7. squared Euclidean distortion is used.

Then

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{k}{2} \log \frac{1}{\varepsilon} + O(1).}$$

Consequently,

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = k.}$$

Proof. Fix y and q. For a small perturbation Δq,

$$g(y, q + \Delta q) - g(y, q) = J_q g(y, q) \Delta q + o(\|\Delta q\|).$$

Hence the local squared reconstruction distortion is

$$\|g(y, q + \Delta q) - g(y, q)\|^2 = \Delta q^\top G(y, q) \Delta q + o(\|\Delta q\|^2),$$

where

$$\boxed{G(y, q) = J_q g(y, q)^\top J_q g(y, q).}$$

Full column rank implies that $G(y, q)$ is positive definite.

Thus, locally, the inverse reconstruction problem is equivalent to a k-dimensional quadratic rate-distortion problem under the positive-definite metric $G(y, q)$. A local linear transformation diagonalizing G reduces the problem to ordinary Euclidean quadratic distortion. High-resolution rate-distortion theory then gives

$$R_{\mathrm{IC}}(\varepsilon) = \frac{k}{2} \log \frac{1}{\varepsilon} + O(1).$$

The $O(1)$ term incorporates the conditional density and the local metric determinant but does not affect the coefficient of $\log(1/\varepsilon)$.

Therefore

$$\mathfrak{d}_{\mathrm{IC}} = 2 \lim_{\varepsilon \downarrow 0} \frac{\frac{k}{2} \log(1/\varepsilon) + O(1)}{\log(1/\varepsilon)} = k.$$

$\square$

20.7 Local inverse geometry

The metric

$$G(y,q) = J_q g(y,q)^\top J_q g(y,q)$$

contains more information than the leading rate coefficient. It measures how errors in the quotient coordinates are amplified when translated into errors in the reconstructed causal variable.

Under the assumptions of Theorem 20.6, let

$$0 < \lambda_{\min}(y,q) \le \lambda_{\max}(y,q)$$

denote the smallest and largest eigenvalues of $G(y,q)$. Then locally,

$$\lambda_{\min}(y,q)\|\Delta q\|^2 \lesssim \|g(y,q+\Delta q) - g(y,q)\|^2 \lesssim \lambda_{\max}(y,q)\|\Delta q\|^2.$$

Thus the condition number

$$\kappa(y,q) = \frac{\lambda_{\max}(y,q)}{\lambda_{\min}(y,q)}$$

measures local anisotropy of the inverse reconstruction problem.

Proof. Since $G(y,q)$ is positive definite,

$$\lambda_{\min}\|\Delta q\|^2 \le \Delta q^\top G \Delta q \le \lambda_{\max}\|\Delta q\|^2.$$

The first-order expansion in the proof of Theorem 20.6 gives

$$\|g(y,q+\Delta q) - g(y,q)\|^2 = \Delta q^\top G \Delta q + o(\|\Delta q\|^2),$$

which yields the stated local inequalities for sufficiently small Δq. □

The inverse-causal information dimension and the inverse conditioning of the mechanism measure different aspects of difficulty. The former determines the leading logarithmic growth of the information rate; the latter determines how efficiently small quotient errors translate into causal reconstruction errors. A problem may therefore have low inverse information dimension but severe local ill-conditioning.

20.8 A representative finite constant

Under stronger global regularity assumptions, the $O(1)$ term in Theorem 20.6 can be refined.

Suppose that $Q \mid Y$ has a sufficiently regular conditional density and that the local metric determinant is integrable. Define

$$G(Y,Q) = J_q g(Y,Q)^\top J_q g(Y,Q).$$

Under the corresponding high-resolution regularity conditions, the rate admits the representative expansion

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{k}{2}\log\frac{1}{\varepsilon} + C_{\mathrm{IC}} + o(1),}$$

where

$$\boxed{C_{\mathrm{IC}} = h(Q \mid Y) + \frac{1}{2}E[\log\det G(Y,Q)] + \frac{k}{2}\log\frac{k}{2\pi e}.}$$

The expression for C_{IC} should be understood as a representative high-resolution constant under the stronger regularity and distortion assumptions under which the corresponding conditional high-resolution formula holds. Unlike the coefficient $k/2$, the constant term is sensitive to the precise quantization, distortion, support, boundary, and conditional-density assumptions.

20.9 Manifold-valued causal quotients

The ambient dimension of a representation need not equal its intrinsic inverse information dimension.

Suppose that, conditional on $Y = y$, the unresolved quotient states lie on a smooth d-dimensional manifold

$$\mathcal{M}_y \subseteq \mathbb{R}^m,$$

with

$$d \leq m.$$

Let

$$\psi_y : V_y \subseteq \mathbb{R}^d \to \mathcal{M}_y$$

be a local coordinate chart and define

$$\widetilde{g}(y, \theta) = g(y, \psi_y(\theta)).$$

The induced inverse metric is

$$\boxed{G_{\mathcal{M}}(y, \theta) = J_\theta \widetilde{g}(y, \theta)^\top J_\theta \widetilde{g}(y, \theta).}$$

[Manifold Inverse Rate Theorem] Suppose the conditional quotient law is regular on $\mathcal{M}_y$, and suppose $\widetilde{g}(y, \cdot)$ is locally bi-Lipschitz with full-rank Jacobian in local coordinates. Then

$$\boxed{R_{\mathrm{IC}}(\varepsilon) = \frac{d}{2} \log \frac{1}{\varepsilon} + O(1).}$$

Consequently,

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = d.}$$

Proof. In a local chart,

$$Q = \psi_y(\Theta), \qquad \Theta \in \mathbb{R}^d.$$

The inverse reconstruction becomes

$$X = \widetilde{g}(Y, \Theta).$$

By assumption, the map from Θ to X is locally bi-Lipschitz. The local squared distortion is therefore equivalent to a positive-definite quadratic form in the d coordinates of Θ:

$$\|\widetilde{g}(y, \theta + \Delta\theta) - \widetilde{g}(y, \theta)\|^2 = \Delta\theta^\top G_{\mathcal{M}}(y, \theta) \Delta\theta + o(\|\Delta\theta\|^2).$$

High-resolution rate-distortion theory in the d local coordinates therefore gives

$$R_{\mathrm{IC}}(\varepsilon) = \frac{d}{2} \log \frac{1}{\varepsilon} + O(1).$$

Since the leading coefficient is invariant under smooth local coordinate changes, the result is intrinsic to the manifold and does not depend on its ambient representation. $\square$

20.10 Connection to causal revelation

The preceding results combine directly with the geometric-information theorem of Appendix 18.

[Inverse-Causal Complexity Decomposition] Suppose the assumptions of Theorem 18.9 hold and that the regular rate-dimension bridge applies. Then

$$\boxed{\mathfrak{d}_{\mathrm{IC}} = d_I(Q \mid Y) = k - r_{\mathrm{rev}}.}$$

If

$$d_I(Q) = k,$$

then

$$\boxed{d_I(Q) = r_{\mathrm{rev}} + \mathfrak{d}_{\mathrm{IC}}.}$$

Proof. By Theorem **??**,

$$\mathfrak{d}_{\mathrm{IC}} = d_I(Q \mid Y).$$

By Theorem 18.9,

$$d_I(Q \mid Y) = k - r_{\mathrm{rev}}.$$

Combining the two equalities gives

$$\mathfrak{d}_{\mathrm{IC}} = k - r_{\mathrm{rev}}.$$

If

$$d_I(Q) = k,$$

then

$$d_I(Q) = r_{\mathrm{rev}} + \mathfrak{d}_{\mathrm{IC}}.$$

□

20.11 Interpretation

The preceding theorem establishes the three-level decomposition

$$\boxed{\begin{aligned} \text{causal latent complexity} &= d_I(Q), \\ \text{causal revelation} &= r_{\mathrm{rev}}, \\ \text{inverse information burden} &= \mathfrak{d}_{\mathrm{IC}}. \end{aligned}}$$

Under the regular full-dimensional setting,

$$\boxed{d_I(Q) = r_{\mathrm{rev}} + \mathfrak{d}_{\mathrm{IC}}.}$$

The interpretation is therefore not that every latent coordinate must be transmitted to reverse the causal mechanism. Rather, only the unresolved portion of the *causal quotient* must be supplied. The original latent variable U may contain arbitrary representational redundancy, and the observed outcome may reveal some of the causally relevant quotient. The inverse-causal rate measures what remains.

The resulting hierarchy is

$$\boxed{U \longrightarrow Q \longrightarrow Q \mid Y \longrightarrow R_{\mathrm{IC}}(\varepsilon).}$$

The first transition removes causal redundancy. The second captures observational revelation. The third quantifies the information required to resolve the remaining causal uncertainty at a specified reconstruction accuracy.

This completes the information-theoretic bridge from the structural causal quotient to the operational difficulty of inverse reconstruction.

21 Appendix F

22 Additional Examples and Technical Lemmas

This appendix collects additional examples and technical lemmas that illustrate the general theory and provide reusable results for applications. The examples emphasize four distinctions developed throughout the paper: (i) latent representation versus causal quotient, (ii) structural invertibility versus observational revelation, (iii) local quotient dimension versus global multiplicity, and (iv) causal information loss versus inverse information burden.

22.1 Additive disturbances and complete causal relevance

Consider

$$Y = h(X) + U,$$

where $h : \mathcal{X} \to \mathcal{Y}$ is injective.

For fixed u, the response function is

$$f_u(x) = h(x) + u.$$

Suppose

$$f_u = f_{u'}$$

for all x. Then

$$h(x) + u = h(x) + u'$$

for all x, and hence

$$u = u'.$$

Thus the response-function map

$$\Phi_f(u) = f_u$$

is injective.

For the additive mechanism

$$Y = h(X) + U,$$

the causal quotient is isomorphic to U:

$$Q \cong U.$$

In particular,

$$\sigma(Q) = \sigma(U).$$

Proof. Injectivity of Φ_f implies that the structural equivalence classes are singletons. Therefore the quotient map does not collapse any latent states, and Q is measurably equivalent to U. $\square$

If h is invertible, then

$$X = h^{-1}(Y - U).$$

Thus the entire causally relevant disturbance is required for exact structural inversion unless some of it is revealed by Y or auxiliary information.

If X and U are continuously distributed and $Y = h(X) + U$, the quotient dimension is the dimension of U. Yet the observation need not reveal any quotient dimension. For example, if

$$X \in \mathbb{R}^p, \qquad U \in \mathbb{R}^k,$$

and the observation provides only their sum in a compatible k-dimensional representation, then the projected quotient fiber generally remains k-dimensional.

This illustrates that

$$d_I(Q) = k$$

does not imply that the observation reduces the inverse burden. Causal relevance and observational revelation are distinct objects.

22.2 Redundant latent representations

Consider

$$Y = X + U_1 + U_2,$$

where the primitive latent vector is

$$U = (U_1, U_2) \in \mathbb{R}^2.$$

The response function is

$$f_{(u_1,u_2)}(x) = x + u_1 + u_2.$$

Therefore

$$(u_1, u_2) \sim_f (u_1', u_2')$$

if and only if

$$u_1 + u_2 = u_1' + u_2'.$$

For

$$Y = X + U_1 + U_2,$$

the causal quotient is

$$\boxed{Q = U_1 + U_2.}$$

Proof. Two latent states induce the same response function exactly when their sums coincide:

$$x + u_1 + u_2 = x + u_1' + u_2' \quad \forall x.$$

Hence the equivalence classes are indexed by $u_1 + u_2$. $\square$

The primitive latent dimension is two, but the causal latent dimension is one:

$$d_I(U) = 2, \qquad d_I(Q) = 1$$

under ordinary regularity.

This example demonstrates why the primitive latent dimension is not the appropriate measure of causal complexity. The second latent coordinate is not independently relevant to the entire intervention-response function. Only the sum survives the causal quotient.

22.3 Nonlinear redundant representations

Consider

$$Y = X + U_1^2 + U_2^2.$$

Then

$$f_{(u_1,u_2)}(x) = x + u_1^2 + u_2^2.$$

Consequently,

$$(u_1, u_2) \sim_f (u_1', u_2')$$

if and only if

$$u_1^2 + u_2^2 = u_1'^2 + u_2'^2.$$

Thus

$$\boxed{Q = U_1^2 + U_2^2.}$$

The quotient collapses both rotational redundancy and sign redundancy in the primitive representation.

If $U \in \mathbb{R}^2$ has a regular distribution, then Q is scalar. The causal information dimension is therefore one rather than two:

$$d_I(Q) = 1.$$

The example also illustrates why the quotient should be defined through the entire response-function map rather than by inspecting individual coordinates of U. A coordinatewise simplification would not reveal the full structural redundancy.

22.4 A mechanism with partial revelation

Consider

$$X = (X_1, X_2), \qquad Q = (Q_1, Q_2),$$

and

$$Y = \begin{pmatrix} X_1 + Q_1 \\ X_2 \\ Q_2 \end{pmatrix}.$$

The observation directly reveals Q_2 but not Q_1.

The Jacobian with respect to X is

$$D_x F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

while

$$D_q F = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The image of $D_x F$ is

$$\operatorname{Im} D_x F = \operatorname{span}\{(1,0,0), (0,1,0)\}.$$

Modulo this image, only the third component of $D_q F$ contributes an independent direction. Hence

$$r_{\text{rev}} = 1.$$

Since $k = 2$,

$$\dim \mathcal{A}_y = k - r_{\text{rev}} = 1.$$

Indeed, conditioning on $Y = y$ fixes Q_2, while Q_1 can vary provided X_1 changes correspondingly.

Under a regular two-dimensional prior for Q,

$$d_I(Q) = 2, \qquad d_I(Q \mid Y) = 1,$$

and therefore

$$\mathcal{R}_{\text{IC}} = d_I(Q) - d_I(Q \mid Y) = 1.$$

The corresponding high-resolution inverse-causal rate is

$$R_{\text{IC}}(\varepsilon) = \frac{1}{2} \log \frac{1}{\varepsilon} + O(1).$$

22.5 Observation with no quotient revelation

Consider

$$Y = X + Q, \qquad X, Q \in \mathbb{R}^k.$$

Then

$$D_x F = I_k, \qquad D_q F = I_k.$$

Since

$$\operatorname{Im} D_q F \subseteq \operatorname{Im} D_x F,$$

the quotient derivative vanishes:

$$\overline{D}_q F = 0.$$

Hence

$$r_{\mathrm{rev}} = 0.$$

The projected fiber is

$$\mathcal{A}_y = \{q : y - q \in \mathcal{X}\}.$$

When $\mathcal{X}$ contains an open neighborhood of the relevant points, this is locally k-dimensional. Thus

$$d_I(Q \mid Y) = k$$

under regularity.

The observation therefore provides no asymptotic reduction in quotient dimension, even though it obviously contains information about Q in the ordinary mutual-information sense.

This is an important distinction:

$$I(Q; Y) > 0$$

does not imply

$$d_I(Q) - d_I(Q \mid Y) > 0.$$

Ordinary mutual information measures statistical dependence; inverse-causal revelation measures reduction in the intrinsic dimension of the unresolved causal quotient.

22.6 Complete revelation

Consider

$$Y = (X, Q).$$

Then

$$D_x F = \begin{pmatrix} I_p \\ 0 \end{pmatrix}, \qquad D_q F = \begin{pmatrix} 0 \\ I_k \end{pmatrix}.$$

The quotient directions are linearly independent of the outcome directions generated by X, so

$$r_{\text{rev}} = k.$$

Consequently,

$$\dim \mathcal{A}_y = 0.$$

Indeed, conditional on $Y = y = (x, q)$,

$$Q = q$$

is known exactly. Therefore

$$d_I(Q \mid Y) = 0$$

and

$$R_{\text{IC}}(\varepsilon) = O(1)$$

as $\varepsilon \downarrow 0$.

The leading logarithmic inverse information burden vanishes.

22.7 Structural noninvertibility cannot be repaired by information

Consider

$$Y = X^2 + Q, \qquad X \in \mathbb{R}.$$

Even if Q is observed exactly,

$$X^2 = Y - Q.$$

For any positive value of $Y - Q$, both

$$X = \sqrt{Y - Q}$$

and

$$X = -\sqrt{Y - Q}$$

produce the same outcome.

Suppose the conditional support of $X \mid Q = q$ contains both x and $-x$ for a set of q with positive probability. Then no measurable decoder

$$g(Y, Q)$$

can recover X exactly almost surely.

Proof. For any such x,

$$x^2 + q = (-x)^2 + q.$$

Thus the same pair (Y, Q) is compatible with two distinct values of X. A deterministic measurable decoder must assign the same value to both cases and therefore cannot equal X almost surely when both possibilities have positive conditional probability. □

The result emphasizes the logical order of the theory:

$$\boxed{\text{structural invertibility} \quad \text{must precede} \quad \text{information accounting.}}$$

No amount of additional information about the latent quotient can resolve ambiguity generated by a noninjective structural map unless that information also contains information about the antecedent itself.

22.8 Finite global ambiguity

Local dimension does not capture finite multiplicity.

Consider

$$Y = X^2, \qquad X \in \{-1, +1\}.$$

The inverse fiber is

$$\{x : x^2 = y\}.$$

At $y = 1$, there are two possible antecedents:

$$X \in \{-1, +1\}.$$

Yet the inverse set is zero-dimensional.

Suppose that, conditional on $Y = y$, the projected quotient fiber $\mathcal{A}_y$ is finite and nonempty. Then

$$d_I(Q \mid Y = y) = 0$$

under the usual information-dimension definition, regardless of the number of points in $\mathcal{A}_y$.

Proof. A finite-valued random variable has quantization entropy bounded uniformly in the quantization scale:

$$H([Q]_\delta \mid Y = y) \leq \log |\mathcal{A}_y|$$

for sufficiently small δ. Therefore

$$\frac{H([Q]_\delta \mid Y = y)}{\log(1/\delta)} \longrightarrow 0.$$

□

Thus information dimension measures the continuous component of inverse complexity. It deliberately ignores finite discrete multiplicity.

This motivates a possible refinement of the theory in which inverse complexity is decomposed into a continuous dimensional component and a discrete global-identification component. Such a decomposition is not required for the main results.

22.9 A nonlinear partially revealed quotient

Consider

$$Y = \begin{pmatrix} X_1 + Q_1^2 \\ X_2 + Q_2 \\ Q_2 \end{pmatrix},$$

with

$$X, Q \in \mathbb{R}^2.$$

The quotient remains

$$Q = (Q_1, Q_2),$$

because distinct values of Q_1 or Q_2 produce distinct response functions.
The Jacobians are

$$D_x F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

and

$$D_q F = \begin{pmatrix} 2Q_1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Modulo $\operatorname{Im} D_x F$, the first two rows are absorbed by changes in X. The third component remains as an independent quotient direction. Hence

$$r_{\text{rev}} = 1$$

whenever the regularity assumptions hold.
Therefore

$$\dim \mathcal{A}_y = 1.$$

The observation reveals one quotient dimension, namely the component represented directly by Q_2, while one dimension remains unresolved.

22.10 A lemma on quotient maps and measurable decoders

The following elementary result is useful whenever a representation is shown to contain the complete causal quotient.

Let Q and W be random elements taking values in standard Borel spaces. If

$$\sigma(Q) \subseteq \sigma(W) \quad \text{modulo null sets,}$$

then there exists a measurable function r such that

$$Q = r(W) \quad \text{a.s.}$$

Proof. This is the measurable factorization property for standard Borel spaces. Since Q is measurable with respect to $\sigma(W)$ modulo null sets, Q admits a measurable version that factors through W. $\square$

Consequently, if

$$X = g(Y, Q),$$

then

$$X = g(Y, r(W))$$

almost surely. Thus quotient-complete auxiliary information is sufficient for inverse reconstruction whenever the structural mechanism itself is invertible.

22.11 A lemma on bi-Lipschitz invariance

The rate and information-dimension results are invariant under regular changes of quotient coordinates.

Let Q and $\widetilde{Q}$ be random elements in Euclidean spaces and suppose

$$\widetilde{Q} = T(Q),$$

where T is bi-Lipschitz on the relevant support. Then

$$d_I(\widetilde{Q} \mid Y) = d_I(Q \mid Y)$$

whenever the conditional information dimensions exist.

Proof. A bi-Lipschitz map changes distances only by fixed multiplicative factors:

$$c\|q - q'\| \leq \|T(q) - T(q')\| \leq C\|q - q'\|.$$

Consequently, a quantization at scale δ in one coordinate system can be covered by a bounded number of cells at scales proportional to δ in the other. The corresponding conditional quantization entropies therefore differ by at most $O(1)$ under the regularity assumptions. Dividing by $\log(1/\delta)$ gives the same limiting information dimension. $\square$

This establishes that the inverse information dimension is an intrinsic property of the unresolved quotient rather than an artifact of the particular coordinates used to represent it.

22.12 A lemma on the quotient derivative

The following identity makes explicit the connection between the quotient derivative and ordinary Jacobian rank.

Let

$$F : \mathcal{X} \times \mathcal{Q} \to \mathcal{Y}$$

be differentiable at (x, q). Then

$$\boxed{\operatorname{rank} \overline{D}_q F = \operatorname{rank} DF - \operatorname{rank} D_x F.}$$

Proof. The image of the full derivative is

$$\operatorname{Im} DF = \operatorname{Im} D_x F + \operatorname{Im} D_q F.$$

The quotient derivative is precisely the composition

$$T_q\mathcal{Q} \xrightarrow{D_q F} T_y\mathcal{Y} \longrightarrow T_y\mathcal{Y} / \operatorname{Im} D_x F.$$

Its image is therefore

$$\frac{\operatorname{Im} D_x F + \operatorname{Im} D_q F}{\operatorname{Im} D_x F}.$$

The dimension of this quotient is

$$\dim \operatorname{Im} DF - \dim \operatorname{Im} D_x F,$$

which gives

$$\operatorname{rank} \overline{D}_q F = \operatorname{rank} DF - \operatorname{rank} D_x F.$$

□

22.13 A lemma on the projected fiber

Let

$$\mathcal{F}_y = F^{-1}(y)$$

and

$$\mathcal{A}_y = \pi_Q(\mathcal{F}_y).$$

At every regular point $(x, q) \in \mathcal{F}_y$,

$$D\pi_Q \left(T_{(x,q)} \mathcal{F}_y \right) = \ker \overline{D}_q F.$$

Proof. Since

$$T_{(x,q)} \mathcal{F}_y = \ker DF,$$

a vector v_q belongs to the image of the restricted projection if and only if there exists v_x such that

$$(v_x, v_q) \in \ker DF.$$

Equivalently,

$$D_x F v_x + D_q F v_q = 0.$$

This holds precisely when

$$D_q F v_q \in \operatorname{Im} D_x F,$$

which is equivalent to

$$\overline{D}_q F(v_q) = 0.$$

Therefore

$$D\pi_Q(\ker DF) = \ker \overline{D}_q F.$$

□

22.14 A local dimension conservation identity

The preceding lemmas yield an immediate local conservation law.

Under the regularity assumptions of Appendix **??**,

$$\boxed{k = r_{\mathrm{rev}} + \dim \mathcal{A}_y.}$$

Proof. By the Projected-Fiber Dimension Theorem,

$$\dim \mathcal{A}_y = k - r_{\mathrm{rev}}.$$

Rearranging gives

$$k = r_{\mathrm{rev}} + \dim \mathcal{A}_y.$$

□

The identity has a direct interpretation. Every local quotient direction is either revealed by the observation or remains in the projected inverse fiber:

$$\boxed{\text{quotient directions} = \text{revealed directions} + \text{unresolved directions.}}$$

22.15 Discrete analogue of dimension conservation

For discrete variables, the analogous decomposition is expressed through conditional entropy rather than information dimension.

If Q is discrete, then

$$H(Q) = I(Q; Y) + H(Q \mid Y).$$

The continuous-dimensional identity

$$d_I(Q) = \mathcal{R}_{\mathrm{IC}} + d_I(Q \mid Y)$$

therefore has the same formal structure as the ordinary entropy decomposition, but with intrinsic dimensional complexity replacing entropy.

The analogy should not be interpreted as an equality between the two theories. Entropy is sensitive to probability masses and, in continuous settings, depends on coordinate-scale conventions through differential entropy. Information dimension instead captures the asymptotic number of degrees of freedom required to resolve a variable at arbitrarily fine resolution.

22.16 A lemma on zero-dimensional inverse burden

Suppose that for almost every y, the conditional projected quotient fiber $\mathcal{A}_y$ is finite and the conditional quotient law is supported entirely on that fiber. Then

$$d_I(Q \mid Y) = 0.$$

Under the regular rate-dimension bridge,

$$\mathfrak{d}_{\mathrm{IC}} = 0.$$

Proof. The first statement follows from Proposition 22.8. Applying Theorem **??** gives

$$\mathfrak{d}_{\mathrm{IC}} = d_I(Q \mid Y) = 0.$$

□

The conclusion does not imply unique inverse identification. A finite set of unresolved alternatives has zero information dimension but may require a nonzero finite number of bits to distinguish.

22.17 Summary of the examples

The examples establish the following hierarchy.

Mechanism	**Inverse structure**	**Main lesson**
$Y = h(X) + U$	$Q = U$	No causal compression
$Y = X + U_1 + U_2$	$Q = U_1 + U_2$	Latent redundancy
$Y = X + U_1^2 + U_2^2$	$Q = U_1^2 + U_2^2$	Nonlinear redundancy
$Y = X + Q$	$r_{\mathrm{rev}} = 0$	No dimensional revelation
$Y = (X, Q)$	$r_{\mathrm{rev}} = k$	Complete revelation
$Y = (X_1 + Q_1, X_2, Q_2)$	$r_{\mathrm{rev}} = 1$	Partial revelation
$Y = X^2 + Q$	Noninvertible	Information cannot fix structural ambiguity
Finite inverse fiber	$d_I(Q \mid Y) = 0$	Finite ambiguity is dimensionless

Taken together, these examples reinforce the central sequence of the theory:

$$\boxed{\begin{array}{c}\text{primitive latent state} \longrightarrow \text{causal quotient} \longrightarrow \text{projected inverse fiber} \\ \longrightarrow \text{conditional information dimension} \longrightarrow \text{inverse-causal rate}\end{array}}$$

The response-function quotient determines which latent distinctions matter causally. The projected fiber determines which of those distinctions remain unresolved after observing the outcome. Conditional information dimension measures the intrinsic size of that unresolved set, while inverse-causal rate-distortion converts that size into an operational information requirement.

This sequence also makes clear why no single conventional object captures the entire problem. A latent representation can be redundant, an observation can statistically depend on the quotient without reducing its intrinsic dimension, and a structurally noninvertible mechanism can remain impossible to reverse even with complete knowledge of the causally relevant latent state.