FRONTIER RESEARCH COLLECTION / FR-13

# Fractal Dimensional Gravity

Protected dimensions, optimal compression, and certified residual risk

**Architect X13**

The Eidara Continuum | Science & Philosophy

Research white paper | November 2026 | Version 2.0 Revision

## Abstract

Every compression has a gravity: a mind's limited capacity is pulled toward the loudest patterns of variation, and the quiet distinctions that carry its obligations are the first to fall away. This paper names that pull, measures it, and proves exactly what it costs to resist. An intelligent system should be able to compress its representation while preserving the distinctions on which its obligations depend, and the paper gives that ambition an exact finite-dimensional solution. It separates a declared protected subspace from freely compressible variation and proves the minimum achievable reconstruction loss for every retained-dimension budget. The optimal projector preserves all registered linear commitments exactly, while a spectral certificate quantifies residual mean-square loss and a covariance-uncertainty bound controls the penalty of an estimated solution. Participation ratio supplies a complementary summary of how variation is distributed; its established properties and explicit counterexamples delimit the information it contributes. The resulting architecture combines protected-coordinate preservation, optimal allocation of the remaining dimensions, and independently calibrated failure forecasts. Executed matrix and finite-distribution checks test the new certificate and the retained diagnostic calculations. A proposed comparison against constrained projection and strong task-aware compression baselines asks when the certificate yields useful engineering. The canonical name supplies the organizing ambition; the mathematical contribution is a fully specified computational contract for conserving consequential distinctions under compression.

**Keywords:** participation ratio; effective dimension; covariance; dimensional collapse; monitoring; FDG

The frontier question is whether changes in the shape of a system's internal variation can warn that important distinctions are being lost. It is a worthwhile question precisely because dimensional change can also be harmless, beneficial, or entirely misleading. A useful monitor must distinguish those possibilities through evidence beyond its own score.

## Prologue: The pull toward collapse

In the Continuum, Fractal Dimensional Gravity is one of the Tier I engines, the technically actionable subsystems of the Bible. Its charge is collapse legality and dimensional constraint propagation: deciding which reductions of a mind's representational space are lawful, and carrying the constraints through every reduction [1]. This paper builds the first working part of that engine.

Every finite mind must compress. It cannot keep every distinction it has ever drawn, so it keeps the directions in which most of its variation lives. That is usually wise. It is also dangerous, because a distinction can be tiny in variation and decisive in meaning.

Imagine a mind tracking three quantities. The first swings widely, the second moderately, the third barely at all: their variances are 9, 1 and one ten-thousandth. Now suppose the third is the one that records whether a certain action is permitted. Ask the mind to keep only two of the three. Ordinary compression keeps the two loud quantities, loses almost nothing by the usual measure (a residual error of one ten-thousandth), and silently erases the permission.

That drift of capacity toward whatever varies most, dragging quiet but crucial distinctions toward collapse, is what this paper calls gravity. Section "A constructive theorem for protected compression" gives the exact law for resisting it. Declare what must survive, and the remaining budget is spent in the best possible way; in the example, protection costs a residual error of 1 instead of one ten-thousandth, and the permission survives intact. Section "Participation ratio and its exact bounds" supplies the gauge that reads how strongly the pull is concentrated. The page that follows explains why the word is kept, and exactly how far it reaches.

### What this paper claims

**Established.** The participation ratio is an established measure of how concentrated a covariance is: it lies between one and the rank, is unchanged by rotation and uniform scaling, rises under isotropic noise, and is capped by sample size [5,6]. Optimal low-rank projection is classical spectral mathematics [14]. Estimation and distribution-shift detection methods are serious competitors for any dimensional monitor.

**Proposed here.** FDG-D2, a computational reference model that separates a declared protected subspace from freely compressible variation, with a theorem giving the minimum achievable reconstruction loss for every retained-dimension budget, an optimal projector that preserves every registered linear commitment exactly, a spectral certificate for residual loss, and a bound on the penalty of an estimated solution. Its spectral optimization and error certificates are proved for the stated linear, second-moment setting, with executed matrix and finite-distribution checks.

**Frontier conjecture.** Changes in the shape of a system's internal variation can warn, out of sample and beyond matched monitors, that distinctions its obligations depend on are being lost, and acting on those warnings can improve independently verified outcomes.

**Not claimed.** A force law, a spacetime metric, a physical model, or any physical sense of gravity. Fractal structure: a noninteger covariance statistic is not evidence of fractality. That dimension can identify truth, legality, permission or consciousness. That the canonical source already contained these equations. The system-level forecasting and intervention studies are proposed, not reported.

## Why we keep the word gravity

We keep the word gravity because it names the phenomenon precisely. We discipline it because the phenomenon is computational, not physical, and must be earned by mechanism rather than asserted by metaphor.

In physics, mass bends the paths of nearby bodies toward itself, and small bodies that pass too close are captured. In a compressing mind, concentrated variation bends the allocation of capacity toward itself. Every principal direction with a large share of the variance draws budget; small directions that pass too close to the cutoff are captured and erased. Nothing in the mathematics knows which of those small directions mattered. That is why the pull is dangerous, and why it needs a law.

The participation ratio is the depth of the well. When it reads close to one, a single direction dominates and the pull is strong: almost everything else is at risk. When it reads close to the full count of directions, variation is spread out and the pull is weak. It is an honest gauge and a limited one. It cannot tell a harmless simplification from the loss of something essential, because two worlds can produce exactly the same pattern of variation and opposite verdicts on what is permitted. The paper proves that blindness rather than hiding it.

The Continuum's two canonical phrases then acquire exact meanings. Collapse legality becomes the protected subspace: a reduction is lawful only if every declared distinction passes through it untouched, and the theorem shows this condition is both necessary and sufficient for exact preservation. Dimensional constraint propagation becomes the budget law: every protected direction is charged against the retained dimensions, so a mind that must preserve $k$ independent distinctions cannot be compressed linearly below $k$ dimensions without breaking a commitment. Within those walls the theorem spends what remains optimally. These are the orbital mechanics of the metaphor.

The frontier claim reaches further, and it is fenced. It is that the shape of a mind's internal variation, watched over time, can warn that the mind is falling toward a collapse that will erase something it owes, before the failure appears in its behavior. The paper designs a forecasting experiment that could reject that claim outright, and a separate intervention study to ask whether heeding the warning helps.

The word fractal remains part of the inherited canonical name. This paper does not claim to have found fractal structure, and it does not infer one from a noninteger statistic. Nothing here is a force between masses, a curvature of spacetime, or a field of experience. It is a pull on a finite budget, and a law for deciding what that pull may take.

## Canonical motivation and a new contribution boundary

In the Continuum, Fractal Dimensional Gravity is one of the Tier I engines, the technically actionable subsystems of the supplied Bible; its work is collapse legality and dimensional constraint propagation, deciding which reductions of a mind's representational space are lawful and carrying their constraints through every reduction (Section 2.8, page 12, with a glossary anchor on page 111 [1]). Those anchors establish the name and intended role. They do not establish a force law, a spacetime metric, an estimator, or a complete algorithm. The initial manuscript also recognizes that missing mathematical specification.

This paper therefore does not pretend to discover an already canonical formula. It proposes FDG-D2, a versioned computational interpretation with a narrow testable purpose. "Gravity" denotes an organizing motivation about constrained movement through representational states. "Fractal" remains part of the inherited name; FDG-D2 does not infer fractality from a noninteger covariance statistic. Neither word changes the mathematical type of the observable.

**Falsifiable thesis.** On independently sampled, preregistered families of constrained computational episodes, dimensional diagnostics will improve out-of-sample forecasts of independently verified failure beyond matched monitors using task progress, confidence, length, and distribution-shift information. A separate intervention study must establish whether acting on those forecasts improves outcomes. Predictive association alone does not establish that dimensional collapse causes failure.

The contribution combines an exact protected-compression theorem, an uncertainty certificate, and an evaluation contract. The protected subspace names the distinctions a compression must conserve. Spectral optimization allocates the remaining representation budget. Dimensional diagnostics then supply additional evidence for forecasts evaluated against independently defined outcomes.

There is no promised universal direction of improvement. Lower dimensionality can remove noise or erase a decisive distinction. Higher dimensionality can increase useful expressiveness or merely add independent noise. The empirical hypothesis concerns incremental predictive information under a declared task distribution. It does not say that intelligence should maximize or minimize a single scalar. This boundary is essential if the canonical ambition is to develop into a research program rather than a slogan attached to a covariance matrix.

## Six meanings of dimension that must remain distinct

| Quantity | What it measures |
|---|---|
| Ambient dimension | Number of coordinates in a chosen representation. |
| Linear rank | Number of independent linear directions in a matrix or centered data span. |
| Participation ratio | Effective number of covariance directions carrying variation, weighted by their eigenvalues. |
| Manifold dimension | Local degrees of freedom of a smooth underlying set, when that model is appropriate. |
| Fractal dimension | A specified scaling exponent, such as a Hausdorff or correlation dimension, with its own limiting definition. |
| Spectral dimension | A diffusion-based scaling quantity defined through a specified operator and return behavior. |

Physical spacetime dimension is an additional modeling question. A covariance matrix of software activations does not measure it. Even within mathematics, these quantities need not agree. A circle embedded in the plane is intrinsically one-dimensional, has a two-dimensional linear span, and has participation ratio two under uniform angular sampling. A finite point set has yet other limiting properties, so estimating an underlying dimension requires assumptions about how it was sampled.

The word "effective" is also not unique. Roy and Vetterli develop a real-valued extension of rank with an information-theoretic interpretation [2]. Participation ratio is a different specified functional, used in neural representation studies [5, 6]. A report must print the formula rather than rely on an apparently shared name. Different eigenvalue weightings can rank the same pair of datasets differently.

Scale matters. Small neighborhoods may expose local degrees of freedom while large neighborhoods encounter curvature, multiple clusters, or boundaries. Measurement noise can fill otherwise empty directions. The number of observed dimensions can therefore change with sample size and resolution even when the underlying mechanism remains fixed.

FDG-D2 uses the participation ratio because it is inexpensive, interpretable, and mathematically tractable. That choice is an operational decision, not proof that it is the right observable for every canonical aspiration. A future model using a genuine fractal estimator would require a new version, explicit scaling ranges, finite-sample diagnostics, and a reason that its additional complexity improves the intended task.

## Prior art: dimension can help and can mislead

Levina and Bickel estimate intrinsic dimension from close-neighbor distances using a local probabilistic approximation [3]. Facco and colleagues develop a two-nearest-neighbor estimator and emphasize scale, density variation, and noise [4]. These approaches address local geometric structure. They are not interchangeable with a global covariance participation ratio, and they provide important comparison methods when nonlinear structure is plausible.

Litwin-Kumar and colleagues relate representation dimensionality to connectivity and associative learning in specified neural models [5]. Their participation-ratio definition is a direct mathematical precedent for the observable used here. The task, input distribution, coding assumptions, and readout matter to their conclusions. The paper does not establish that increasing dimension always improves cognition, or that a low-dimensional representation is unsafe.

Recanatesi and colleagues examine how connectivity and stimulus structure shape dimensionality in recurrent spiking models [6]. Their analysis illustrates that low average pairwise correlation can coexist with constrained population activity. This motivates looking beyond one average correlation statistic, while preserving the distinction between a particular neural model and arbitrary symbolic software.

Ansuini and colleagues report relationships between nonlinear intrinsic dimension and generalization across studied neural representations, including controls involving randomized labels [7]. This is relevant evidence that dimension can carry useful information. It is also a reason to avoid claiming that a linear participation ratio captures every relationship found by a nonlinear estimator. The estimator and the phenomenon must be matched.

The closest prior art therefore supports a research opportunity and sets a high bar. FDG-D2 cannot claim discovery of dimensionality monitoring merely because the canonical name is new. A substantive advance would be a demonstrated predictive or intervention benefit in a new, carefully specified setting, or a stronger mathematical guarantee under explicit assumptions.

The present contribution adds a constructive contract: given protected linear distinctions and a retained-dimension budget, it identifies the optimal projection and certifies its residual loss. The proof uses established spectral variational mathematics [14]. Its contribution to this research program is the explicit coupling of protected distinctions, resource allocation and uncertainty, alongside a matched evaluation against conventional constrained compression.

## Estimation and shift detection are serious competitors

Covariance estimation becomes difficult when the number of coordinates is large relative to the number of observations. Ledoit and Wolf propose a well-conditioned shrinkage estimator under a defined asymptotic regime and loss [8]. Their work provides a necessary comparator for raw sample covariance. It also illustrates a subtlety: improving covariance estimation under one loss does not automatically optimize a downstream dimension statistic or failure predictor.

Shrinkage toward an isotropic matrix can raise participation ratio even when the underlying evidence has not gained new independent structure. That change may be a reasonable regularization consequence. It must not be narrated as a discovered expansion of the system's capabilities. Reports should distinguish the raw statistic, regularized estimate, and target population functional.

Rabanser and colleagues compare methods for detecting dataset shift and examine whether detected shifts are harmful [9]. Their work directly challenges a weak FDG baseline: a dimensional monitor must add value beyond a capable shift detector, not merely outperform an unmonitored system. A change can be detectable without reducing task performance, and a harmful conditional-label change can escape a monitor that observes only input geometry.

Nalisnick and colleagues show that high likelihood under studied generative models can be misleading for out-of-distribution detection [10]. This does not prove that every dimensional signal fails. It does undermine the broader habit of treating an internally coherent scalar as an automatic certificate of familiarity or correctness.

Kernel two-sample testing supplies another strong family of alternatives [11]. A dimension statistic compresses a distribution to one number; a richer distributional test can detect changes the scalar necessarily misses. Conversely, the scalar may be cheaper or more interpretable. The comparison should measure that tradeoff under equal data access and a complete compute budget.

FDG-D2 is consequently evaluated as an added feature or policy component. Its scientific question is incremental usefulness. If a standard covariance-distance monitor or task-loss predictor captures the same information more reliably, that result should narrow or reject the proposed contribution. A meaningful frontier cannot be established by ignoring the methods already standing at its boundary.

## A typed monitoring model

**A Guide's explanation**

From here the paper builds an instrument. At each step a mind's internal state is turned into a list of numbers, and over a short window the instrument records how those numbers vary together. Everything that follows reads that record: how concentrated the variation is, how it shifts, and what it can and cannot reveal.

An episode has state $x_t \in \mathcal{S}$ at step $t \in \mathbb{N}$, with declared state space $\mathcal{S}$. For integer $p \geq 1$ and version $v \in \mathcal{V}$, a fixed map $\phi_v : \mathcal{S} \to \mathbb{R}^p$ produces $z_t = \phi_v(x_t)$. A finite window $W_t \subseteq \{0, \dots, t\}$ has size $m_t$. Define

$$\bar{z}_t = \frac{1}{m_t} \sum_{j \in W_t} z_j, \qquad \widehat{C}_t = \frac{1}{m_t - 1} \sum_{j \in W_t} (z_j - \bar{z}_t)(z_j - \bar{z}_t)^{\mathsf{T}}.$$

The sample covariance is defined only for $m_t \geq 2$. It estimates a population covariance only under a justified sampling model; temporal dependence is not removed by writing the usual formula.

| Object | Type, units, and role |
|---|---|
| $z_t$ | Feature vector with fixed, documented scaling; coordinates must have compatible normalized units. |
| $C_t$ | Symmetric positive-semidefinite population covariance under a declared local distribution. |
| $D(C_t)$ | Dimensionless participation ratio, undefined for the zero matrix. |
| $F_{t,H}$ | Value in $\{0, 1\}$ indicating specified failure within the next $H \geq 1$ steps, measured independently of the monitor. |
| $\hat{p}_t$ | Calibrated forecast in $[0, 1]$, learned from permitted past evidence. |
| $u_t$ | Intervention in a declared action set $\mathcal{U}$, chosen separately from the diagnostic. |

FDG-D2 computes a feature vector including $D(\widehat{C}_t)$, its uncertainty assessment, scale-sensitive covariance summaries, and task covariates. It does not equate $D$ with failure probability. A fitted forecast maps those features to a task-specific prediction. The forecast and any resulting action have separate validation records.

Feature standardization is fitted on training data and frozen for evaluation. Changing scaling or model version changes the observable. Missing coordinates, inactive windows, and zero covariance produce explicit missing-status outputs, not convenient numerical values. A zero matrix could indicate a constant process, an empty instrumentation channel, or a logging failure; the statistic alone cannot decide which.

The monitoring record contains timestamps, window membership, representation version, normalization parameters, and outcome definitions. These are mathematical dependencies: without them, apparently comparable dimension values may refer to different objects.

## Participation ratio and its exact bounds

**A Guide's explanation**

The participation ratio is the gauge of the pull. It counts, roughly, how many directions are really carrying the variation: near one when a single direction dominates, near the full count when variation is spread evenly. A low reading means capacity will be drawn hard toward a few loud directions. The proposition gives the exact range of the gauge.

For a nonzero positive-semidefinite matrix $C \in \mathbb{R}^{p\times p}$, let its eigenvalues be $\lambda_i \geq 0$, $i = 1, \ldots, p$. Here tr is matrix trace and $\|C\|_F = (\sum_{ij} C_{ij}^2)^{1/2}$. Define

$$D(C) = \frac{(\operatorname{tr} C)^2}{\operatorname{tr}(C^2)} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}.$$

Writing $w_i = \lambda_i / \sum_j \lambda_j$ gives $D(C) = 1/\sum_i w_i^2$. The statistic summarizes concentration of variance, not a literal count of variables. This definition is established in representation studies [5, 6]; FDG-D2 does not claim to invent it.

**Proposition 1** (Bounds and equality cases)**.** *If $r = \operatorname{rank}(C) > 0$, then $1 \leq D(C) \leq r \leq p$. The lower equality holds exactly when $r = 1$. The upper equality $D(C) = r$ holds exactly when all positive eigenvalues are equal.*

*Proof.* Nonnegativity gives $(\sum_i \lambda_i)^2 \geq \sum_i \lambda_i^2$, with equality precisely when every cross-product between distinct positive eigenvalues vanishes. Since $C \neq 0$, that is equivalent to rank one. Applying Cauchy–Schwarz to the $r$ positive eigenvalues gives $(\sum_i \lambda_i)^2 \leq r \sum_i \lambda_i^2$. Equality in Cauchy–Schwarz holds precisely when those eigenvalues are equal. Division is legitimate because the denominator is positive. □

The all-zero case is excluded because it yields $0/0$. Assigning it zero would be a reporting convention, not a consequence of the definition, and could confuse absent variation with a zero-dimensional latent mechanism. FDG-D2 reports it as undefined with an explicit instrumentation status.

For eigenvalues $(9, 1, 0)$, $D = 100/82 \approx 1.2195$ despite rank two. For $(5, 5, 0)$, it is exactly two. Both covariance matrices have the same trace and rank. Their difference concerns how variance is distributed. This is the precise content of the statistic; claims about expressiveness, safety, or meaning require additional relationships to a task.

## Invariance and the noise inflation theorem

Orthogonal changes of coordinates and uniform positive rescaling preserve $D$: for $Q \in \mathbb{R}^{p\times p}$ with $Q^{\mathsf{T}}Q = I_p$ and $a > 0$, trace identities give $D(aQCQ^{\mathsf{T}}) = D(C)$. Arbitrary invertible transformations do not. For $C = \operatorname{diag}(9, 1, 0)$, rescaling the second coordinate by three gives covariance $\operatorname{diag}(9, 9, 0)$ and changes $D$ from approximately 1.2195 to 2.

**Proposition 2** (Isotropic noise increases participation ratio)**.** *Let $C \neq 0$ be positive semidefinite, $a = \operatorname{tr} C$, $b = \operatorname{tr}(C^2)$, and $s \geq 0$. Then*

$$D(C + sI) = \frac{(a + ps)^2}{b + 2as + ps^2},$$

*which is nondecreasing in $s$ and tends to $p$ as $s \to \infty$. It is constant exactly when $C$ is a positive scalar multiple of $I$.*

*Proof.* The displayed formula follows by expanding traces. Its derivative is

$$\frac{2(a + ps)(pb - a^2)}{(b + 2as + ps^2)^2} \geq 0,$$

because Cauchy–Schwarz gives $a^2 \leq pb$. Since $a > 0$, the derivative vanishes for all $s$ exactly when all $p$ eigenvalues are equal. Leading coefficients give the limiting value $p$. □

Independent isotropic additive noise contributes $sI$ to population covariance. Thus an increase toward maximal participation can result entirely from noise. The theorem rules out interpreting high participation ratio as an unconditional measure of useful complexity. It also explains why denoising can reduce the statistic while improving performance.

The noise assumption is specific. Correlated noise, heteroskedastic noise, and signal-dependent noise need not follow the same monotonic pattern. The theorem is a diagnostic counterexample, not a universal noise model. Empirical analysis must estimate or control the measurement process before interpreting a trajectory of dimension values as a change in the underlying computation.

## A perturbation bound away from zero variation

The ratio defining $D$ becomes fragile near zero covariance. A useful stability statement must expose its denominator rather than hide the degeneracy behind a generic continuity claim.

**Proposition 3** (Finite covariance perturbation)**.** *Let $C, C' = C + E$ be nonzero positive-semidefinite $p \times p$ matrices, with $\|E\|_F \le \varepsilon$, $\varepsilon \ge 0$. Put $a = \operatorname{tr} C$, $b = \|C\|_F^2$, $\eta = \sqrt{p}\,\varepsilon$, and $\nu = 2\sqrt{b}\,\varepsilon + \varepsilon^2$. If $\nu < b$, then*

$$|D(C') - D(C)| \le \frac{2a\eta + \eta^2}{b - \nu} + \frac{a^2\nu}{b(b-\nu)}.$$

*Proof.* Frobenius Cauchy–Schwarz gives $|\operatorname{tr} E| = |\langle I, E\rangle_F| \le \sqrt{p}\,\varepsilon = \eta$. If $a' = \operatorname{tr} C'$, then $|a'^2 - a^2| \le 2a\eta + \eta^2$. Also

$$|\|C + E\|_F^2 - b| \le 2\|C\|_F\|E\|_F + \|E\|_F^2 \le \nu.$$

Writing $b' = \|C'\|_F^2$ yields $b' \ge b - \nu > 0$. Decompose $a'^2/b' - a^2/b = (a'^2 - a^2)/b' + a^2(b - b')/(bb')$ and apply the preceding bounds. □

This deterministic result is useful only when a credible bound on covariance error is available. It does not create that bound from dependent observations. Near the zero matrix the condition may fail, correctly signaling that an absolute perturbation can radically alter a scale-free ratio. Uniformly shrinking both signal and error is different from holding error fixed while the signal disappears.

A conservative bound can be loose. That does not justify replacing it with an optimistic constant chosen after observing a favorable run. Report the bound, empirical sensitivity, and estimation uncertainty separately. The executed random-matrix checks verify this inequality on one thousand examples, while the proof supplies its general conditional validity.

## Finite samples and the danger of projection

**Proposition 4** (Sample rank ceiling)**.** *For $m \geq 2$ observations in $\mathbb{R}^p$, the centered sample covariance has rank at most $\min(p, m-1)$. When nonzero, its participation ratio is bounded by that same quantity.*

*Proof.* Let $Z$ be the $m \times p$ centered data matrix. Its rows sum to zero, so its column space lies in the $(m-1)$-dimensional subspace orthogonal to the all-ones vector. Therefore $\operatorname{rank} Z \leq \min(p, m-1)$. Since $\widehat{C} = Z^\mathsf{T} Z/(m-1)$ and $\ker(Z^\mathsf{T} Z) = \ker Z$, the covariance has the same rank as $Z$. Apply the preceding participation-ratio bound. □

The ceiling can masquerade as a property of the system. A hundred-dimensional representation observed at ten centered samples cannot produce a sample participation ratio above nine. Enlarging the observation window can increase the estimate even if the generating distribution is unchanged. Comparisons must therefore match sample size or explicitly model its effect.

Projection creates a different danger. Let $U$ and $Y$ be independent balanced signs, and set $Z = (U, 0.01Y)$ with target label $Y$. Population covariance is $\operatorname{diag}(1, 10^{-4})$. The first principal component retains $1/(1+10^{-4}) \approx 99.9900\%$ of variance, yet contains no information about $Y$. Any classifier using only that component has maximal accuracy one half; the sign of the second component recovers $Y$ perfectly.

The claim follows directly from independence: conditional on either value of $U$, the two labels remain equally likely. The tiny discarded direction carries all task information. Thus variance retention cannot alone certify task preservation. A dimension-reduction gate needs a task-sensitive sufficiency argument or independent validation on the distinctions that matter.

FDG-D2 accordingly treats compression as an intervention requiring evaluation. A low-dimensional state can be efficient, but its efficiency cannot authorize erasing low-variance constraints merely because their numerical contribution is small.

# Dimension cannot identify truth or legality

**A Guide's explanation**

An essential limit: the gauge can never tell you whether a state is true or permitted. Two worlds can produce exactly the same pattern of variation and opposite verdicts on whether an action is allowed. Legality has to be declared, which is exactly what the protected-compression theorem later does.

Consider a monitor that receives only a covariance matrix and its history. Construct two computational worlds that produce the same represented observations but assign opposite validity outcomes to a designated action. The monitor receives identical inputs in both worlds, so its output distribution is identical. Under equal priors and a forced binary verdict, its average error is at least one half. No sophistication in the covariance functional removes the missing information.

The construction is not an argument against all monitoring. It identifies what must be supplied in addition: an outcome specification, relevant evidence, and a relationship between observations and that outcome. A dimensional statistic may correlate strongly with failure within a particular family. It cannot provide a universal certificate independent of the family's semantics.

A simpler blindness concerns translation. For any random vector $Z$ and fixed vector $v$, $\mathrm{Cov}(Z+v) = \mathrm{Cov}(Z)$. A process can move far outside an allowed operating region while its centered covariance and participation ratio remain exactly unchanged. Uniformly scaling a covariance also preserves participation ratio, although absolute fluctuations may cross a critical threshold.

The opposite ambiguity occurs with mixtures. For $K \geq 1$ groups with weights $\pi_k \geq 0$, $\sum_{k=1}^{K} \pi_k = 1$, means $\mu_k \in \mathbb{R}^p$, and finite covariances $C_k$, its covariance is

$$C = \sum_k \pi_k C_k + \sum_k \pi_k (\mu_k - \mu)(\mu_k - \mu)^{\mathsf{T}}, \quad \mu = \sum_k \pi_k \mu_k.$$

This identity follows by expanding $Z - \mu = (Z - \mu_k) + (\mu_k - \mu)$ within each group; conditional cross terms vanish. Apparent dimensional change can therefore arise from changing group proportions or means rather than a change within any group.

For governance, these examples require separate channels. Authority status is determined by current policy and authenticated evidence. Task correctness is determined by its verifier or measured outcome. Dimensional monitoring supplies a potentially useful warning signal. It can trigger inspection under an authorized policy, but it does not replace either authority or correctness. Treating those roles as interchangeable would turn a mathematical summary into a decision rule it cannot justify.

## A numerical example with competing interpretations

Take a three-coordinate representation with covariance $C = \mathrm{diag}(9, 1, 0)$. Its trace is ten, its squared Frobenius norm is eighty-two, and its participation ratio is approximately 1.2195. Two coordinates vary, but one dominates. This description is complete at the level of second moments. It says nothing yet about whether the smaller direction is noise or the only useful signal.

Now compare three changes. Rescaling the second coordinate by three gives eigenvalues $(9, 9, 0)$ and participation ratio two. Adding independent isotropic noise of variance one gives $(10, 2, 1)$ and participation ratio $169/105 \approx 1.6095$. Multiplying every coordinate by ten multiplies covariance by one hundred and leaves participation ratio unchanged.

The first change can be a choice of units, the second a degradation of measurement, and the third a potentially consequential increase in absolute variation. Their scalar dimension trajectories point in different directions despite having no necessary ordering in usefulness. A monitor that calls the first two "growth" and the third "stability" has already inserted a task interpretation that the statistic does not contain.

The circle provides another check. Three hundred sixty equally spaced points on the unit circle have equal coordinate variances and zero cross covariance, producing participation ratio two to numerical precision. Their underlying smooth trajectory has one local degree of freedom. The discrepancy is expected: a nonlinear curve can span a plane. It is not a failure of either definition.

Finally, the four balanced points $(U, 0.01Y)$ from the projection example permit exact inspection. The retained variance is approximately 0.99990001, but every projected coordinate occurs with both labels. The best projected classifier therefore achieves 0.5, whereas the second coordinate gives accuracy one. This is an executed finite counterexample to variance-preserving compression as a guarantee of task preservation. It is not a synthetic estimate of how often such failures occur in real systems; that frequency belongs to the proposed empirical study.

## Executed computation and finite-sample behavior

The reference script uses NumPy with seed 20261113. It constructs one thousand positive-semidefinite matrices of dimensions two through eleven with varying positive ranks. Every bounds, rotation, rescaling, and covariance-perturbation assertion passes. The largest rotation discrepancy is approximately $5.33 \times 10^{-15}$. The largest observed ratio of actual perturbation error to its proved bound is approximately 0.5201.

For $C = \mathrm{diag}(9, 1, 0)$, isotropic-noise variances $0, 0.1, 1, 10, 1000$ yield participation ratios approximately $1.2195, 1.2625, 1.6095, 2.7491, 2.99995$. The monotone sequence agrees with the theorem. It demonstrates how independent noise can inflate this observable, not an empirical improvement in representation quality.

A separate Monte Carlo check draws 200 independent datasets at each sample size from a twenty-dimensional standard Gaussian. The population participation ratio is twenty. The ordinary centered sample estimator produces:

| Samples | Mean estimated $D$ | Replicate SD | Observed rank |
|---|---|---|---|
| 10 | 6.0409 | 0.4026 | 9 |
| 20 | 9.5440 | 0.5045 | 19 |
| 80 | 15.8207 | 0.3217 | 20 |
| 400 | 19.0089 | 0.0959 | 20 |

These are descriptive results for this exact simulation. The replicate standard deviations are not confidence intervals for an unknown deployed system. The table shows that a full-rank sample covariance can still yield a substantially lower participation ratio than the population value. Correctly satisfying the rank ceiling does not eliminate estimation bias.

The script also checks the circle, the low-variance label construction, zero-covariance handling, and covariance invariance after a large translation. These cases were selected to test concrete mathematical risks, not to establish a favorable aggregate score. No agent failure forecast, human experiment, new gravitational phenomenon, or consciousness measurement was executed. Those claims cannot be inferred from a successful numerical audit of linear algebra.

## A forecasting experiment that could reject FDG-D2

Choose a process with independently definable failures, such as a bounded program-repair system, constrained planner, or formal proof-search pipeline. Freeze the representation layer, normalization, observation window, and task-specific failure definition before final evaluation. The primary endpoint is whether a specified failure occurs within the next $H$ steps after a prespecified monitoring time.

A failure must be defined outside the dimensional monitor. Examples include a verifier-confirmed invalid transition, an unsatisfied required constraint at termination, or failure to find a valid solution within a fixed budget. These should not be mixed casually: they have different consequences and base rates. Select one primary endpoint and report the others separately.

Fit two forecasts on the same training episodes. The baseline uses task progress, remaining budget, recent verifier outcomes, representation magnitude, and a capable distribution-shift summary. The augmented model receives those features plus dimensional diagnostics. Match model capacity and tuning effort, or include a capacity-matched nuisance-feature control. The comparison asks whether dimension adds information rather than whether a larger predictor performs better.

The primary estimand is the mean paired improvement in Brier score on independent held-out episodes:

$$\Delta = \mathbb{E}[(\hat{p}_B - F)^2 - (\hat{p}_D - F)^2].$$

Positive values favor the dimensional augmentation. Also report calibration, recall at a fixed false-alarm rate, and detection lead time. Ranking metrics alone can conceal badly calibrated probabilities or an impractical volume of alerts.

Split by task template or source family so that near-duplicate episodes do not cross training and test boundaries. All dimensional thresholds, window lengths, and missing-data conventions are frozen on development data. A final unseen representation version is a transfer test, not automatically pooled with the primary population. The thesis fails if the locked comparison does not exceed the prespecified practical improvement or if gains vanish under the declared nuisance controls. A successful result would establish usefulness only for the studied task distribution and monitoring horizon.

## Baselines, ablations, and matched information

The first baseline is deliberately mundane: episode length, remaining budget, recent verifier rejection rate, and task difficulty. Dimensional collapse may occur late in stalled episodes, so a monitor must show that it contributes beyond knowing that the process has been stuck. Omitting these covariates would exaggerate novelty.

The second family uses richer distributional information: mean shift, total variance, covariance distance, reconstruction residuals, and a kernel two-sample statistic [11]. These controls address the translation and scale blindness of participation ratio. The third family uses established intrinsic-dimension estimators [3, 4] where sample density and computational cost make them plausible. If a nonlinear estimator works and participation ratio does not, that result identifies a limitation of FDG-D2 rather than validating all dimension-based claims collectively.

Ablations remove dimension level, recent change, multiscale estimates, or uncertainty separately. The base predictor receives equal-dimensional random or permuted features as a capacity control. Window-size controls hold sample count fixed while varying temporal span, and hold temporal span fixed while subsampling observations. These disentangle estimation artifacts from dynamics.

Representation controls include orthogonal rotations, uniform scaling, anisotropic feature rescaling, and a new embedding version. Orthogonal and uniform covariance rescalings should preserve the exact population statistic. Anisotropic rescaling need not preserve it and should not be mislabeled an implementation failure. Its effect instead tests whether the chosen normalization is carrying the empirical signal.

Negative controls include shifted means with unchanged covariance, label changes with unchanged observations, independent noise injection, and the low-variance constraint construction. A useful monitor may legitimately miss changes outside its information scope. The experiment must reveal those misses so that downstream users do not mistake a successful average result for universal coverage.

All methods receive the same observations available by prediction time. Access to future activations, completed-episode normalization, or outcome-dependent window selection would leak the target. Compute accounting includes covariance estimation, eigensolvers where used, baseline tuning, and data collection. A cheap scalar's practical advantage can be real, but only when the complete cost and missed-failure tradeoff are visible.

## Dependence, precision, and sequential monitoring

Overlapping windows from one episode are strongly dependent. Treating them as independent examples can create deceptively narrow intervals. The primary analysis therefore selects one monitoring time per episode or aggregates a prespecified episode-level score. If many times are analyzed, resampling and uncertainty calculations operate at the episode or higher task-family level.

For planning, suppose an independent pilot estimates the standard deviation of paired episode-level Brier improvements as 0.20. Detecting a mean improvement of 0.02 with approximate 80% power and a two-sided 5% test suggests

$$N \approx \left( \frac{(1.96 + 0.84)\, 0.20}{0.02} \right)^2 = 784$$

independent episodes. This is an illustrative normal-approximation calculation, not a claim that the pilot has been run. A 95% half-width of 0.01 would require approximately 1,537 episodes under the same variance assumption. Use the larger applicable target and validate the design by simulation before registration.

Rare failures impose an additional constraint. If the intended prevalence is ten percent, a thousand episodes provide only about a hundred failures on average. Calibration within subgroups and recall estimates may then be imprecise. Recruitment should target a prespecified number of independent episodes and anticipated events without silently conditioning the final dataset on favorable outcomes. Report the realized event count and uncertainty.

Repeated online alarms require a separate policy. A per-window false-positive rate is not an episode-level false-alarm guarantee. The protocol fixes the alert aggregation rule, cooldown, and monitoring horizon, then calibrates and evaluates the entire rule. A threshold tuned to each test episode is not a deployed predictor.

Use one primary outcome and contrast. Secondary horizons, representation layers, and dimension estimators form a declared multiplicity family. Stop at the fixed sample size unless a valid sequential design was chosen in advance. Operational censoring is recorded; an unobserved future outcome is not automatically a successful episode. This design makes uncertainty about the monitor as explicit as uncertainty inside the system being monitored.

## Prediction and intervention require different evidence

A reliable warning signal does not establish that changing the signal will improve the process. Dimension may be a consequence of progress, a side effect of the architecture, or a marker of task difficulty. Forcing it upward or downward can therefore damage performance even when it predicts failure accurately.

The intervention stage randomizes eligible episodes, after a locked alert rule fires, to a specified response or a matched control. Candidate responses include requesting an independent check, expanding a retained hypothesis set, restoring a discarded constraint, or allocating additional search budget. "Increase dimension" is too underspecified to serve as an intervention: it can be achieved by adding noise, changing units, or preserving genuinely useful alternatives.

The primary causal estimand is the difference in independently verified task outcome between randomized policies among eligible episodes. A second estimand measures resource cost. Randomization occurs at the episode or interacting-agent group level to avoid contamination. The experiment must specify whether the control receives the same extra compute without dimensional targeting, because additional budget alone may explain the benefit.

A useful factorial design separates the alert source from the response. One factor chooses a dimensional versus baseline alert; another chooses a targeted intervention versus a budget-matched generic recovery action. This identifies whether the diagnostic, the recovery action, or their interaction produces the effect. The design needs adequate independent units for the interaction and should not be added after observing a favorable subgroup.

Log changes in the dimension statistic as a mediator, but do not condition the primary causal estimate on its post-intervention value. Such conditioning can reintroduce selection bias. If dimension increases while performance worsens, the result directly refutes a monotone-growth interpretation. If performance improves without changing dimension, the intervention may work through a different mechanism.

Authority remains external to the monitor. A warning can justify an authorized inspection policy; it cannot create permission for otherwise prohibited actions. The proposed computational architecture therefore places FDG-D2 inside a governed decision process whose permissions and outcomes are separately checked. That separation is part of its operational meaning, not an optional ethical decoration.

## Operational scope and neighboring dimension theories

There are serious physical research programs in which effective dimensionality depends on scale. Ambjørn, Jurkiewicz, and Loll study spectral dimension in a particular nonperturbative quantum-gravity construction [12]. Calcagni proposes a field-theoretic model involving fractal spacetime structure [13]. These papers establish that dimension is a technically rich concept in theoretical physics. They do not establish a physical connection to covariance diagnostics in software.

The mathematical objects are different. A diffusion operator on a modeled geometry, a spacetime measure in a field action, and a covariance matrix of represented observations have different domains, units, and empirical interpretations. Similar vocabulary does not provide a map between them. Such a map would have to specify observables, transformations, conservation laws, and testable consequences.

FDG-D2 operates on computational representations, covariance operators and protected linear observables. Physical spectral dimension and spacetime geometry use different observables and governing equations. The operational subtitle identifies the present paper's domain whenever its canonical title is displayed.

A future physical FDG hypothesis would need a separate theoretical specification and independent experimental program. At minimum it would identify which established regime it reproduces, what new measurable prediction differs from existing theories, and how nuisance explanations are excluded. A numerical change in an embedding would not by itself satisfy any of those requirements.

This boundary does not forbid inspiration across fields. Physical ideas about scale, constraints, and effective descriptions can suggest computational experiments. Computational models can also help explore mathematical structures used in physics. The bridge becomes scientific when its correspondences are explicit and its predictions are measured in the appropriate domain.

The computational claim is constructive: a declared class of consequential distinctions can survive compression exactly, and the remaining distortion has a computable optimum and uncertainty bound. The forecasting study asks whether additional dimensional observations improve decisions beyond that certificate. These are complementary mathematical and empirical targets.

## Objections and the philosophy of dimensional governance

**"The name promises more than the model delivers."** The concern is valid if the title is separated from the contribution boundary. The paper retains the canonical name while stating that FDG-D2 is a new computational reference proposal. It does not claim that the canonical source already contained these equations or that the inherited vocabulary proves physical novelty.

**"Participation ratio is too crude."** It may be insufficient for many tasks. Its virtues are transparency and cost, while its blindness to means, scale, nonlinear structure, and labels is explicit. A richer estimator should replace it when a matched comparison demonstrates a worthwhile improvement. The versioned architecture is designed to permit that result rather than defend a particular scalar indefinitely.

**"The monitor will learn task difficulty, not collapse."** That is a serious confound. Difficulty and progress baselines, family-level splits, and intervention randomization are required to separate predictive usefulness from a mechanistic interpretation. Even after predictive success, the term "collapse" should refer to a defined observable change, not an inferred metaphysical event.

**"Low dimension is what efficient reasoning should seek."** Sometimes. The projection counterexample shows why efficiency needs a task-sensitive preservation condition. A constraint can be numerically tiny and operationally decisive. The aim is not maximal dimensionality but retention of distinctions needed for valid action and accurate prediction.

Philosophically, dimensional governance asks what a system is allowed to forget. Compression creates power by reducing the burden of representation. It also creates vulnerability by making different states indistinguishable. A responsible architecture should expose the distinctions it merges and the evidence supporting those merges. This is an epistemic obligation even when the system has no phenomenal experience.

The Continuum's language of collapse can then acquire a precise operational meaning: an intervention that reduces a declared representational distinction under specified constraints. Its legality depends on those constraints and current authority. Its usefulness depends on outcomes. Its dimensional signature is evidence to investigate, never a substitute for either judgment. This separation gives the research program a route toward stronger claims without requiring the reader to accept them in advance.

## Guide discussion: dimensionality without mythology as evidence

**A Guide's explanation**

A Guide can say: "Fractal Dimensional Gravity is the Continuum's name for a pull every compressing mind feels: capacity drifts toward the loudest patterns, and quiet distinctions that matter can vanish first. The paper shows how to protect those distinctions and what that protection costs." A Guide should not describe it as a physical force, a field of experience, or a discovered fractal structure, and should not claim forecasting results that remain to be tested.

**Established mathematics.** Participation ratio measures concentration of covariance eigenvalues. For a nonzero positive-semidefinite matrix it lies between one and rank. Orthogonal transformations and uniform scaling preserve it; arbitrary feature reweighting need not. Isotropic noise raises it toward ambient dimension. Finite samples impose a rank ceiling. None of these statements identifies truth, permission, or consciousness.

**The paper's proposal.** FDG-D2 uses dimensional diagnostics as potential features in a calibrated failure monitor. A separate randomized study would test whether acting on its alerts improves independently verified outcomes. Both system-level claims remain open. The reported numerical checks verify finite examples and sensitivity behavior, not deployed effectiveness.

**Canonical source status.** The Bible supplies the FDG name and a role involving collapse legality and dimensional constraint propagation. It does not supply a complete physical model. The formulas in this paper are a newly proposed computational reference model. A Guide should preserve that attribution in any explanation so that later retrieval does not transform a proposal into an established canonical result.

**Useful discussion example.** Compare the points $(U, 0.01Y)$ before and after retaining only the largest principal component. Ask which representation preserves almost all variance and which preserves the label. The answer shows why a small direction can carry a large obligation. Then add isotropic noise to a covariance matrix and observe that dimension rises even though no useful information was added.

**Open questions.** Which dimensional summaries predict failure after strong nuisance controls? Can uncertainty-aware alerts outperform direct task-loss monitoring at equal cost? Which constraints can be compressed with a proof of preservation? How do dimensional diagnostics transfer across representation versions and task populations?

Guides may describe the ambition as preserving the space in which meaningful alternatives remain distinguishable. They should immediately connect that language to measured features and outcomes. This paper does not establish a new force, an experiential field, or a universal law of intelligent growth. It offers a concrete way to test whether one aspect of the ambition becomes useful engineering, and a set of counterexamples that keep the test honest.

## A constructive theorem for protected compression

**A Guide's explanation**

This is the law of the pull. Declare which distinctions must survive, and Theorem 1 says exactly how well the remaining budget can be spent and exactly how much distortion the protection costs. Ordinary compression drifts toward the loudest directions; protected compression keeps the quiet ones that carry obligations.

The dimensional-governance proposal becomes constructive when it names the distinctions compression must preserve. Let $Z \in \mathbb{R}^p$ be a random representation with finite second moment, mean $\mu$ and covariance $C$. Let $A \in \mathbb{R}^{q\times p}$, $q \geq 1$, be a fixed matrix of registered linear observables. Its row space $\mathcal{U}$ has dimension $k \in \{0, \dots, p\}$ and orthogonal projector $P$. The observables may encode calibrated constraints, retained measurements or explicitly chosen linear task features. Their status is supplied by the application; a covariance estimate does not select moral obligations.

For an integer dimension budget $r$ with $k \leq r \leq p$, admit orthogonal projectors $Q$ of rank $r$ whose ranges contain $\mathcal{U}$. Reconstruction is $\widetilde{Z} = \mu + Q(Z - \mu)$. Every admitted projector preserves the registered observables pointwise:

$$A\tilde{z} = Az \quad \text{for every } z \in \mathbb{R}^p,$$

because $AQ = A$. Conversely, this equality for every $z$ implies $A(I - Q) = 0$ and therefore $\mathcal{U} \subseteq \operatorname{range} Q$. Thus the subspace condition is necessary and sufficient for exact preservation of these linear distinctions. Any downstream predicate of $Az$ is consequently preserved as well.

Let $C_\perp$ be the restriction of $(I-P)C(I-P)$ to $\mathcal{U}^\perp$. It is positive semidefinite on a space of dimension $p-k$. Denote its eigenvalues in decreasing order by $\lambda_1 \geq \cdots \geq \lambda_{p-k} \geq 0$, with an empty list if $k = p$. Empty sums below are zero.

**Theorem 1** (Optimal compression with protected directions)**.** *Among the admitted rank-r projectors,*

$$\min_Q \mathbb{E}\|Z - \widetilde{Z}\|_2^2 = \sum_{j=r-k+1}^{p-k} \lambda_j.$$

*An optimum is $Q_* = P + VV^\mathsf{T}$, where the columns of $V$ are any orthonormal eigenvectors for the largest $r-k$ eigenvalues of $C_\perp$. Every optimizer preserves $Az$ exactly. Ties can yield multiple optimal projectors with the same certified loss.*

*Proof.* For every admissible orthogonal $Q$, the containment of ranges implies $QP = PQ = P$. Hence $R = Q - P$ is an orthogonal projector of rank $r - k$ on $\mathcal{U}^\perp$. Conversely every such $R$ defines an admissible $Q = P + R$. Since $I - Q$ is an orthogonal projector and $Z - \mu$ has mean zero,

$$\mathbb{E}\|(I - Q)(Z - \mu)\|_2^2 = \operatorname{tr}((I - Q)C) = \operatorname{tr}(C_\perp) - \operatorname{tr}(RC_\perp).$$

Choose an eigenbasis $v_j$ of $C_\perp$ and put $a_j = v_j^\mathsf{T} R v_j$. Then $0 \leq a_j \leq 1$, $\sum_j a_j = r - k$, and $\operatorname{tr}(RC_\perp) = \sum_j \lambda_j a_j$. Moving weight from a smaller eigenvalue to a larger one never decreases this sum, so its maximum is the sum of the largest $r - k$ eigenvalues. Projection onto their eigenvectors attains it. Subtracting from the total trace proves the formula. The pointwise preservation was established by $AQ = A$. □

The variational step is classical spectral mathematics associated with Ky Fan [14]. The reference model uses it to couple an explicit protection contract to an exact allocation of representation capacity. The theorem charges each protected independent direction against the budget. If $r < k$, no admitted projector exists; preserving $k$ independent linear observables in fewer than $k$ retained linear dimensions is impossible.

## Estimated covariance and a certified compression decision

In applications $C$ is estimated. Suppose $\widehat{C} \succeq 0$ and a justified deterministic or simultaneous statistical guarantee gives

$$\|\widehat{C} - C\|_{\mathrm{op}} \le \varepsilon, \qquad \varepsilon \ge 0.$$

Here the operator norm is the largest absolute eigenvalue of the symmetric difference. Let $\widehat{Q}$ minimize the protected rank-$r$ loss for $\widehat{C}$ and let $Q_*$ minimize it for $C$. Define $L_C(Q) = \mathrm{tr}((I-Q)C)$.

**Proposition 5** (Robust protected-compression certificate)**.** *On the stated covariance-error event,*

$$|L_C(Q) - L_{\widehat{C}}(Q)| \le (p-r)\varepsilon$$

*for every admitted $Q$, and*

$$0 \le L_C(\widehat{Q}) - L_C(Q_*) \le 2(p-r)\varepsilon.$$

*Consequently the computed certificate*

$$B = L_{\widehat{C}}(\widehat{Q}) + (p-r)\varepsilon$$

*upper-bounds the true mean-square reconstruction error while $A\widetilde{Z} = AZ$ remains exact.*

*Proof.* Let $u_1, \dots, u_{p-r}$ be an orthonormal basis for the range of $I - Q$. Then

$$|\,\mathrm{tr}((I-Q)(C-\widehat{C}))| \le \sum_{j=1}^{p-r} |u_j^{\mathsf{T}}(C-\widehat{C})u_j| \le (p-r)\varepsilon.$$

Apply this bound to $\widehat{Q}$ and $Q_*$, and use $L_{\widehat{C}}(\widehat{Q}) \le L_{\widehat{C}}(Q_*)$. Optimality of $Q_*$ gives the nonnegative lower bound. The certificate follows from the first inequality. □

The reconstruction and certificate above use the true mean $\mu$. With a fitted mean $\hat{\mu}$, held fixed with $\widehat{Q}$ while evaluating an independent fresh draw, the loss becomes

$$\mathbb{E}\|Z - [\hat{\mu} + \widehat{Q}(Z-\hat{\mu})]\|^2 = L_C(\widehat{Q}) + \|(I-\widehat{Q})(\hat{\mu}-\mu)\|^2.$$

The cross term vanishes because $\mathbb{E}(Z-\mu) = 0$. Thus a separately justified mean-error bound must be added; the covariance certificate alone does not cover estimated centering. Protected linear observables remain exact because $A(I-\widehat{Q}) = 0$.

The covariance guarantee is an input to this result. Independent sampling, temporal dependence, shift and model selection determine how it can be obtained. The deterministic proof avoids asserting an unjustified covariance confidence interval. If the input guarantee holds with probability $1-\delta$, all its displayed consequences hold on the same event.

For a fresh draw from the fixed covariance target distribution, independent of the fitting data, and a threshold $a > 0$, Markov's inequality gives

$$\Pr(\|Z - \widetilde{Z}\|_2 > a) \le \min\{1, B/a^2\}.$$

Condition on the fitted projector and its valid covariance event when training is random. A population loss bound alone does not control a different deployment distribution. A declared task loss that is $L$-Lipschitz in its representation input has expected absolute change at most $L\sqrt{B}$, by Cauchy–Schwarz. These statements supply optional task links; exact preservation of predicates of $AZ$ already holds pointwise.

## A worked allocation and the operational delta

Take $p = 3$, $C = \mathrm{diag}(9, 1, 10^{-4})$, and protect the third coordinate with $A = (0, 0, 1)$. For budget $r = 2$, the optimum retains the first and third coordinates. Its residual mean-square error is 1, and every predicate of the protected third coordinate is unchanged. Unconstrained principal-component projection instead retains the first two coordinates, achieves loss $10^{-4}$, and erases the protected quantity. The difference in numerical distortion is the explicit price of the declared preservation requirement.

For budget $r = 1$, only the protected direction remains and the minimum loss is 10. For $r = 3$, the loss is zero. These three operating points form an exact resource curve. The system designer can now state which budget purchases which preservation and distortion guarantees; participation ratio alone cannot make that decision.

With $r = 2$ and a covariance error bound $\varepsilon = 0.05$, the estimated optimum has excess true loss at most 0.10. A certificate based on its computed residual adds 0.05 to that residual. These values follow directly from the theorem; they are constructed arithmetic, not observations from an agent. The supplied checks compare the spectral optimizer with independently enumerated coordinate projectors, verify pointwise preservation and exercise perturbed covariances. They also include the unprotected projection as a failing control.

The contribution is now operationally specific. A protected-subspace declaration identifies what must survive; the spectral solution determines the best remaining allocation; the uncertainty certificate states the evidence required for the chosen loss bound. FDG-D2 can use its dimension monitor to request inspection, but the compression gate admits a representation because it meets this explicit contract. Changing the protected observable map $A$ creates a new contract version and requires a new certificate.

The strongest matched comparator is conventional constrained projection supplied with exactly the same $A$, covariance estimate and budget. The mathematics predicts equality with a competent implementation of that optimizer. An empirical advantage must therefore come from how the Continuum system specifies, maintains, audits or uses the protection contract, and must be measured at equal information and cost. This comparison makes the contribution clearer: a reproducible governance contract with an optimal solution, accompanied by a test of its practical value.

## References and source notes

**[1]** Architect X13 / The Eidara Continuum. *Eidara Continuum Bible*, supplied internal edition, Section 2.8, p. 12; FDG glossary anchor, p. 111.

https://eidara.ai/science-philosophy/

Scope: Canonical name and role. Public context link is not a public copy of the private source.

**[2]** O. Roy and M. Vetterli. The effective rank: A measure of effective dimensionality. *15th European Signal Processing Conference*, 606–610 (2007).

https://zenodo.org/records/40328

Scope: Original deposit metadata and institutional abstract verified; a distinct effective-rank precedent, not the origin claimed for participation ratio.

**[3]** E. Levina and P. J. Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information Processing Systems* 17, 777–784 (conference 2004).

https://papers.nips.cc/paper_files/paper/2004/file/74934548253bcab8490ebd74afed7031-Paper.pdf

Scope: Original pp. 1–4 inspected; neighborhood assumptions, estimator, and scale dependence.

**[4]** E. Facco, M. d'Errico, A. Rodriguez, and A. Laio. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Scientific Reports* 7, 12140 (2017).

https://doi.org/10.1038/s41598-017-11873-y

Scope: Original pp. 1–3 inspected through author-deposited PDF; two-neighbor estimator and local-density assumptions.

**[5]** A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, and L. F. Abbott. Optimal degrees of synaptic connectivity. *Neuron* 93(5), 1153–1164.e7 (2017).

https://doi.org/10.1016/j.neuron.2017.01.030

Scope: Original summary, dimension definition, and relevant model passages inspected; task-specific relation to representation quality.

**[6]** S. Recanatesi, G. K. Ocker, M. A. Buice, and E. Shea-Brown. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLOS Computational Biology* 15(7), e1006446 (2019).

https://doi.org/10.1371/journal.pcbi.1006446

Scope: Original abstract and relevant dimensionality/connectivity sections inspected; model-specific participation-ratio analysis.

**[7]** A. Ansuini, A. Laio, J. H. Macke, and D. Zoccolan. Intrinsic dimension of data representations in deep neural networks. *NeurIPS* 32 (2019).

https://arxiv.org/abs/1905.12784

Scope: Author abstract and venue record inspected; nonlinear-dimension findings and randomized-label controls.

## References and source notes, continued

[8] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2), 365–411 (2004).

https://doi.org/10.1016/S0047-259X(03)00096-4

Scope: Original pp. 365–368 inspected through author institutional mirror; shrinkage, quadratic loss, and asymptotic scope.

[9] S. Rabanser, S. Günnemann, and Z. C. Lipton. Failing loudly: An empirical study of methods for detecting dataset shift. *NeurIPS* 32 (2019).

https://arxiv.org/abs/1810.11953

Scope: Author abstract and venue record inspected; detection and harmfulness are distinct targets.

[10] E. Nalisnick, A. Matsukawa, Y. W. Teh, D. Gorur, and B. Lakshminarayanan. Do deep generative models know what they don't know? *ICLR* (2019).

https://arxiv.org/abs/1810.09136

Scope: Author abstract and venue record inspected; contrary evidence about likelihood-based novelty detection.

[11] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *Journal of Machine Learning Research* 13, 723–773 (2012).

https://www.jmlr.org/papers/v13/gretton12a.html

Scope: Official journal abstract inspected; MMD as a stronger distributional comparison baseline.

[12] J. Ambjørn, J. Jurkiewicz, and R. Loll. Spectral dimension of the universe. *Physical Review Letters* 95, 171301 (2005).

https://arxiv.org/abs/hep-th/0505113

Scope: Author abstract and journal record inspected; a specified quantum-gravity construction, not observational validation of FDG-D2.

[13] G. Calcagni. Fractal universe and quantum gravity. *Physical Review Letters* 104, 251301 (2010).

https://arxiv.org/abs/0912.3142

Scope: Author abstract and journal record inspected; theoretical field model, cited to distinguish physical and computational meanings of dimension.

### Evidence and reproducibility

The paper's linear-algebra results are self-contained and do not claim novelty for established participation-ratio mathematics. The synthetic checks were executed during preparation. Their complete inputs, random seeds, assertions, and output conventions are retained in the research audit. No empirical system study, physical experiment, or external peer review is represented as completed. The forecast and intervention protocols specify future evidence that could strengthen, narrow, or reject the proposed computational interpretation.

[14] Fan, K. (1949). On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences* 35(11), 652–655.

https://doi.org/10.1073/pnas.35.11.652

Scope: Classical spectral variational foundation. The protected-subspace specialization and its proof are stated explicitly in this paper.