Orthogonal metrics for variational inference landscape complexity

The optimization landscape of variational Bayesian inference in sparse regression with correlated features can be rigorously decomposed into two orthogonal complexity dimensions. **Between-basin complexity** captures landscape fragmentation—how many distinct basins exist and the likelihood of converging to an incorrect one—while **within-basin complexity** captures identifiability limitations given convergence to any particular basin. This decomposition maps directly onto the distinction between **algorithm error** (failing to find the right region) and **model specification error** (inherent uncertainty once there).

Existing theoretical frameworks from submodularity theory, spectral analysis, and variational inference convergence provide strong foundations for both metrics, though no single existing quantity perfectly satisfies all the user's desiderata. The key insight is that between-basin complexity should measure properties of the correlation matrix R that create **multiple equivalent or near-equivalent sparse solutions**, while within-basin complexity should measure **local identifiability** within correlation-defined clusters.

Theoretical foundations for between-basin complexity

The between-basin metric should capture landscape fragmentation—specifically, how "blurry" LD structure creates many small basins versus how "clean" block structure creates few large basins. Three theoretical frameworks provide foundations.

Submodularity ratio from Das & Kempe

The **submodularity ratio** γ from Das & Kempe (2011, 2018) directly measures how greedy algorithms can "get stuck" due to interaction effects between variables:

$$\gamma_{U,k}(f) = \min_{L \subseteq U, S: |S| \leq k, S \cap L = \emptyset} \frac{\sum_{x \in S} [f(L \cup \{x\}) - f(L)]}{f(L \cup S) - f(L)}$$

For truly submodular functions γ ≥ 1; for correlated regression objectives, **0 < γ < 1** indicates deviation from submodularity. The crucial theoretical result is that γ provides a lower bound on greedy algorithm approximation quality: forward regression achieves **(1 - e^{-γ}) × OPT**. (jmlr)

The submodularity ratio connects to sparse eigenvalues via **γ_{U,k} ≥ λ_min(R, k + |U|)**, but Das & Kempe show γ can be significantly larger than this eigenvalue bound. (jmlr) This is important because near-singular R matrices may still have large γ when singularities don't align unfavorably with the optimization objective.

However, γ is **NP-hard to compute exactly** and depends on the specific objective function (including the outcome variable), not just R alone. (jmlr) For a metric computable purely from R, we need an R-only proxy.

Restricted eigenvalue conditions and sparse eigenvalues

The **minimum k-sparse eigenvalue** λ_min(R,k) = min_{|S|=k} λ_min(R_S) captures the worst-case conditioning over k-variable subsets. (jmlr) When λ_min(R,k) is small, there exist k-dimensional subspaces where the correlation matrix nearly collapses—creating ridges and valleys in the optimization landscape that partition it into basins.

The **Restricted Eigenvalue (RE) condition** from Bickel, Ritov, and Tsybakov (2009) provides a weaker but more practically relevant condition:

$$\kappa(s, c_0) = \min_{J:|J|\leq s} \min_{\delta\neq 0, \|\delta_{J^c}\|_1 \leq c_0\|\delta_J\|_1} \frac{\|R^{1/2}\delta\|_2}{\|\delta_J\|_2}$$

The RE constant κ → 0 indicates the existence of multiple near-equivalent solutions (between-basin issue), while κ bounded away from zero indicates a unique well-defined basin structure.

Generalized correlation from VI convergence theory

Bhattacharya, Pati, and Yang (2023) introduced **Generalized Correlation (GCorr)** for analyzing coordinate ascent variational inference convergence:

$$\text{GCorr}_\alpha(\pi_n) = \sup_{q_j \in Q_j} \frac{|\Delta_n(q_1, q_2)|}{\sqrt{D_{KL,\alpha}(q_1\|q_1^*) \cdot D_{KL,1-\alpha}(q_2\|q_2^*)}}$$

The key theoretical result: **GCorr < 2 guarantees convergence to a unique global optimum**, while GCorr ≥ 2 indicates potential multimodality. (arXiv) (arxiv) For Gaussian targets, GCorr = 2‖B‖₂ where B is derived from the precision matrix structure. This provides the most direct theoretical link between correlation structure and basin multiplicity in variational inference specifically.

Theoretical foundations for within-basin complexity

The within-basin metric should capture expected identifiability given convergence to some basin—related to credible set size and Fisher information, smoothly handling "soft" blocks.

Effective rank and eigenvalue distribution

The **entropy-based effective rank** (Roy & Vetterli, 2007) captures how many dimensions carry independent information:

$$\text{eRank}(R) = \exp\left(-\sum_i p_i \log p_i\right), \quad p_i = \frac{\lambda_i}{\sum_j \lambda_j}$$

For a correlation matrix, 1 ≤ eRank(R) ≤ p, with lower values indicating more redundancy among predictors. The **Effective Dimensionality Ratio** EDR = eRank(R)/p provides a normalized measure.

For a block of k variables with uniform correlation ρ, the eigenvalues are $\lambda_1 = 1 + (k-1)\rho$ and $\lambda_2 = \ldots = \lambda_k = 1 - \rho$. This yields condition number $\kappa = (1 + (k-1)\rho)/(1 - \rho)$, which demonstrates how "soft" blocks affect identifiability:

ρ	Block κ (k=5)	Effective Rank	Interpretation
0.5	6	~3-4	Moderate correlation, effects partially distinguishable
0.9	46	~1.5	Strong collinearity, only aggregate effect identifiable
0.99	496	~1.0	Near-perfect collinearity, variables interchangeable

Fisher information and credible set theory

The Fisher information matrix for linear regression with correlated design is $\mathbf{I(\beta) = (1/\sigma^2)\, X'X \propto R}$ for standardized predictors. The Cramér-Rao bound $\mathrm{Var}(\hat{\beta}) \geq I(\beta)^{-1} = \sigma^2 R^{-1}$ shows that **credible set volume ∝ det(R)^{-1/2}**.

For Bayesian sparse regression with spike-and-slab priors, Castillo et al. (2015) established that posterior contraction rate is $\sqrt{(s \log p / n)}$ under compatibility conditions, with credible set diameter controlled by the **smallest scaled sparse singular value** and **mutual coherence**. Within a basin, these quantities determine how "fuzzy" the posterior is over nearby solutions.

LD Score analogue for local complexity

From genetics literature, the **LD Score** $\ell_j = \Sigma_k\, r^2_{jk}$ (sum of squared correlations) measures how much each SNP "tags" other variants. A local within-basin metric could average or aggregate LD scores within identified correlation blocks to capture local identifiability.

Constructing orthogonal metrics satisfying the desiderata

Based on the theoretical foundations, here are candidate formulations that maintain orthogonality and satisfy the specified properties.

Metric 1: Between-basin complexity (landscape fragmentation)

Proposed quantity: A transformation of the "effective modularity" of |R| viewed as a weighted adjacency matrix, combined with deviation from clean block structure.

Key desiderata satisfaction:

- *0/1 equivariance*: All R matrices with only 0/1 entries should score identically. This requires the metric to depend on |R| only through quantities invariant to block permutation and size.
- *Continuous/smooth*: Should be differentiable in R entries.
- *Blurry LD scores HIGH*: Correlations near 0.5 create many small "partial" basins.
- *Clean block LD scores LOW*: Correlations near 0 or 1 create few large basins.

Candidate formula:

$$\mathrm{BBC}(R) = \sum_{i<j} h(|r_{ij}|), \quad h(x) = 4x(1-x) \cdot g(x)$$

where h(x) peaks at x = 0.5 and is zero at x = 0 and x = 1, and g(x) is a weighting function. The sum over all pairs captures total "ambiguity" in the correlation structure. This satisfies 0/1 equivariance because h(0) = h(1) = 0.

A more sophisticated version uses the **spectral gap of the thresholded adjacency graph**. Define A_τ with $(A_\tau)_{ij} = \mathbb{1}(|r_{ij}| > \tau)$ and compute the Fiedler eigenvalue (smallest non-zero eigenvalue of the graph Laplacian) across thresholds:

$$\mathrm{BBC}_{\text{spectral}}(R) = -\int_0^1 \lambda_2(L_{A_\tau}) \cdot w(\tau)\, d\tau$$

where w(τ) weights the contribution of each threshold. Higher values indicate more fragmented landscapes.

Alternative based on submodularity proxy: The minimum sparse eigenvalue λ_min(R,k) captures worst-case basin separation. (jmlr) A between-basin score could use:

$$\mathrm{BBC}_{\text{sparse}}(R) = \sum_{k=1}^{k_{\max}} w_k \cdot \left(1 - \frac{\lambda_{\min}(R,k)}{\bar{\lambda}_k}\right)$$

where $\bar{\lambda}_k$ is the expected minimum eigenvalue under independence. This measures deviation from well-posed sparse regression.

Metric 2: Within-basin complexity (local identifiability)

Proposed quantity: A block-averaged effective rank measure that captures expected posterior uncertainty given assignment to any basin.

Key desiderata satisfaction:

- *Smoothly handles soft blocks*: Continuous in correlation values.
- *Captures credible set size*: Related to Fisher information.
- *Basin-agnostic*: Averaged across or expected over possible basins.

Candidate formula: First identify "soft blocks" via spectral clustering on R, then compute within-block effective dimensionality:

$$\mathrm{WBC}(R) = \sum_{b=1}^{B} \frac{|C_b|}{p} \cdot \left(1 - \frac{\mathrm{eRank}(R_{C_b})}{|C_b|}\right)$$

where C_b are the identified clusters and R_{C_b} is the correlation submatrix for cluster b. This averages (1 - EDR) across clusters, weighting by cluster size. Higher values indicate worse identifiability within basins.

Alternative based on condition number: Use block-averaged log condition number:

$$\mathrm{WBC}_{\kappa}(R) = \sum_{b=1}^{B} \frac{|C_b|}{p} \cdot \log(\kappa(R_{C_b}))$$

where κ(R_{C_b}) is the condition number of the within-cluster correlation matrix.

Alternative based on Fisher information: For the full matrix, use the average inverse Fisher information:

$$\mathrm{WBC}_{\mathrm{Fisher}}(R) = \frac{1}{p}\mathrm{tr}(R^{-1})$$

This captures average parameter variance under perfect knowledge of the support, but requires R to be invertible (regularize if needed).

Ensuring orthogonality between the metrics

The key insight for maintaining orthogonality is that **between-basin complexity should depend on the global topology of correlations** (how blocks connect to each other, how "fuzzy" boundaries are), while **within-basin complexity should depend on local geometry** (conditioning within blocks).

The 2×2 decomposition the user specified:

	Low Within-Basin	High Within-Basin
Low Between-Basin	Best case: few basins, each well-identified	Model limitation: few basins but hard to pinpoint within
High Between-Basin	Algorithm limitation: many basins but each clear	Worst case: many basins, each fuzzy

Mathematical separation:

- Between-basin: Measures **inter-block** structure and boundary sharpness
- Within-basin: Measures **intra-block** conditioning and redundancy

Consider two extreme cases:

1. **Perfect block structure with ρ=0.99 within blocks**: BBC = 0 (sharp boundaries), WBC = high (near-singular blocks)
2. **Uniform ρ=0.5 everywhere**: BBC = high (no clear blocks), WBC = moderate (each pair has some distinguishability)

Connections between spectral, information-theoretic, and optimization perspectives

The three perspectives converge on related but distinct aspects of the problem:

Spectral perspective

- **Eigenvalue distribution** of R determines both basin structure (via sparse eigenvalues) and local identifiability (via condition number)
- **Graph Laplacian eigenvalues** of thresholded |R| capture connectivity/fragmentation Csustan
- **Marchenko-Pastur law** provides null distribution for detecting true signal eigenvalues

Information-theoretic perspective

- **Fisher information** $I(\beta) \propto R$ directly determines credible set geometry
- **Mutual information** between predictors and response partitioned across basins
- **Effective degrees of freedom** capture model complexity accounting for correlation

Optimization perspective

- **Submodularity ratio** γ determines greedy algorithm approximation quality (jmlr)
- **GCorr** from VI theory determines CAVI convergence to global vs. local optima
- **Mode connectivity** and barrier heights characterize landscape topology (OpenReview) (UC Berkeley Statistics)

Unifying insight: All three perspectives agree that the eigenstructure of R (and its sparse restrictions) is fundamental. The spectral properties determine both the number of near-equivalent solutions (between-basin, via sparse eigenvalues being small) and the uncertainty within solution basins (within-basin, via condition number). The key to orthogonality is distinguishing **which eigenstructure properties** matter for each metric.

Computational complexity considerations

Quantity	Exact Complexity	Practical Approach
Mutual coherence $\mu = \max$	r_{ij}	
Condition number $\kappa(R)$	$O(p^3)$	Full eigendecomposition
Effective rank eRank(R)	$O(p^3)$	Full eigendecomposition
Sparse eigenvalue $\lambda\_min(R,k)$	NP-hard	Random subset sampling, SDP relaxation
RE constant $\kappa(s,c_0)$	NP-hard	Probabilistic bounds, power iteration
Submodularity ratio γ	NP-hard	Upper/lower bounds from eigenvalues
Graph Laplacian spectrum	$O(p^3)$	Lanczos for largest/smallest eigenvalues
GCorr for Gaussian	$O(p^3)$	Derived from precision matrix

Practical recommendations:

- For $p < 1000$: Full eigendecomposition is feasible; compute condition number, effective rank, spectral gap directly
- For $p > 1000$: Use randomized SVD, Lanczos iteration for extreme eigenvalues, (Wikipedia) random subset sampling for sparse eigenvalue estimates
- For between-basin: Random sampling of k-subsets to estimate $\lambda\_min(R,k)$ distribution
- For within-basin: Block structure identification via spectral clustering, then per-block analysis

---

Existing metrics and their limitations for this application

For between-basin (landscape fragmentation)

LD Score (genetics): $\ell_j = \Sigma_k r^2_{jk}$ measures local correlation burden but doesn't directly capture basin count. Could be transformed via total LD score or LD score variance.

Submodularity ratio γ: Theoretically ideal but depends on the outcome variable, not just R. Das & Kempe's sparse eigenvalue lower bound provides an R-only approximation. (jmlr)

GCorr (VI convergence): Directly measures CAVI convergence difficulty. For Gaussian mean-field, GCorr = $2\|Q_{11}^{-1/2}Q_{12}Q_{22}^{-1/2}\|_2$ depends on how R is partitioned.

Limitation: No existing metric perfectly captures the "blurry LD scores high, clean block scores low" desideratum with 0/1 equivariance.

For within-basin (identifiability)

Condition number κ(R): Standard measure but global, not block-aware.

Effective number of independent tests M_e: From genetics, counts eigenvalues contributing to 99.5% of variance. Related to effective rank.

Irrepresentability gap η: Measures how close the irrepresentability condition is to failing. Continuous in R but support-specific.

Limitation: Most metrics are global; need block-aware versions for within-basin interpretation.

Recommended approach for metric development

Step 1: Compute the full eigendecomposition of R and the graph Laplacian of |R| at multiple thresholds.

Step 2: For between-basin complexity, compute:

- Ambiguity score: $\Sigma_{i<j} 4|r_{ij}|(1-|r_{ij}|)$ (peaks at 0.5, zero at 0/1)
- Fiedler eigenvalue at threshold 0.5 (measures bipartiteness of "ambiguous" correlations) (Csustan)
- Estimated sparse eigenvalue λ_min(R,k) for k = 1, 2, ..., 10 via sampling

Step 3: For within-basin complexity, compute:

- Identify blocks via spectral clustering on |R| with automatic cluster count selection (arXiv)
- Per-block condition number and effective rank
- Weighted average of log(κ) across blocks

Step 4: Validate orthogonality by constructing test cases:

- Block diagonal with varying within-block ρ (should vary WBC, constant BBC)
- Uniform correlation at varying ρ (should vary BBC, moderate WBC)
- Mixed structures with both clear and fuzzy blocks

This framework provides a principled decomposition of optimization landscape complexity into algorithm-relevant (between-basin) and model-relevant (within-basin) components, grounded in established theory from submodularity, variational inference, and spectral analysis.