ChatGPT

Continuous LD-Based Metrics for SuSiE/IBSS Posterior Complexity

Introduction

Fine-mapping methods like SuSiE (Sum of Single Effects) use Bayesian variable selection to identify causal SNPs, but their difficulty can vary greatly with the **linkage disequilibrium (LD)** structure of the genotype matrix X (or its correlation matrix **R**). Highly correlated SNPs can lead to **ambiguous posterior landscapes** with multiple nearly-optimal solutions (posterior *basins of attraction*). This makes inference harder – IBSS (Iterative Bayesian Stepwise Selection), the variational algorithm used by SuSiE, may converge slowly or get trapped in a local optimum of the ELBO [1] [2]. We seek **unsupervised, continuous metrics** derived from X or R (no outcome *y* or model fitting) to predict this difficulty a priori.

We propose two complementary metrics: (1) a **between-basin complexity** metric to estimate how many distinct high-posterior "regions" (modes) might exist due to LD structure, and (2) a **within-basin complexity** metric to quantify the ambiguity or breadth of the posterior within a single mode. These metrics are invariant to SNP labeling (permutation-invariant), continuous in R's entries (no arbitrary thresholds), and provide insight into why certain LD patterns create multiple plausible causal configurations ("blurry" LD) or wide credible sets. We also relate these metrics to existing theory (submodular optimization, variational inference, sparse recovery) and discuss how they could inspire improved optimization strategies (e.g. mirror descent, temperature annealing, multi-basin averaging) for IBSS.

1. Between-Basin Complexity: Distinct Posterior Modes

Definition: *Between-basin complexity* measures the potential number of **distinct high-posterior basins** – essentially, how many different causal configurations could be nearly optimal given the LD in X. Intuitively, if many dissimilar sets of SNPs can explain the data almost equally well (due to correlation structure), the posterior will be **multimodal**, with separate peaks for each configuration. We want a metric that is high in such cases (many modes) and low when the posterior is essentially unimodal. Key requirements: it should reflect *blurry LD* that enables many plausible causal sets, be invariant to permuting SNP indices, treat perfect block-LD structures equitably (any block-diagonal R with entries 0/1 should yield a consistent measure), and vary *continuously* with changes in R (no sharp thresholds).

Candidate Metric – LD Eigenentropy (Effective Rank): One principled way is to measure the **diversity of correlation structure** via the eigenvalues of the LD matrix R. If R has eigenvalues $\lambda_1,\ldots, \lambda_p$ (which sum to p for a correlation matrix), define:

$$H_{\text{spec}} = -\sum_{i=1}^{p} \frac{\lambda_i}{p} \ln\Big(\frac{\lambda_i}{p}\Big),$$

and the **effective rank** $\mathrm{ER}(R) = \exp(H_{\text{spec}})$ [3]. This $\mathrm{ER}(R)$ can be interpreted as the "effective number of independent LD clusters" in the genotype matrix [3]. It ranges from 1 (all SNPs in one perfectly correlated block) to p (all SNPs completely independent). Crucially, it is **permutation-invariant** and continuous in R's entries. In a *perfect block-structured* LD matrix (R entries in ${0,1}$ forming, say, c blocks of equal LD), $\mathrm{ER}(R)=c$ exactly. More generally, if block sizes differ or LD is imperfect, $\mathrm{ER}(R)$ will be a non-integer, reflecting partial fragmentation of LD structure.

- **Rationale:** A high effective rank means the correlation matrix has many non-negligible eigenvalues, i.e. genetic variation spans many roughly independent directions. This suggests multiple distinct combinations of SNPs could contribute to a trait. For example, if SNPs form several moderately correlated groups (no single group dominates), there may be **multiple posterior modes** – one for a causal variant in each group or combination thereof. Conversely, a low effective rank (e.g. one large LD block or highly collinear SNPs) indicates the data lie in essentially one or few dimensions, so the posterior might concentrate in one region (albeit possibly wide). Thus $\mathrm{ER}(R)$ serves as a **continuous proxy for the number of plausible causal basins**.

- **Connections to submodularity:** In feature selection terms, if SNP groups are well-separated (high diversity), selecting one SNP provides diminishing returns for selecting another in the same group – a property akin to submodularity which tends to yield a single global optimum for a greedy algorithm [4] [5]. But if $\mathrm{ER}(R)$ is high, the "greedy" IBSS might face several nearly-optimal choices (one per independent direction). In fact, theoretical work shows that the performance of greedy selection relates to the **smallest sparse eigenvalue** of the covariance matrix [5]. A low sparse eigenvalue (i.e. near-dependencies among features) implies the set function is far from submodular, permitting multiple optima [5]. Our metric $\mathrm{ER}(R)$ is indirectly related: a low minimum eigenvalue (poor conditioning) often accompanies a low effective rank, flagging **multimodal posteriors**. In other words, $\mathrm{ER}(R)$ is high when the LD structure is "spread out" (approximate submodularity holds in each subset), and low when LD creates couplings that violate diminishing returns, potentially yielding multiple modes.

- **Example:** Consider two extreme cases. *(a)* **Independent SNPs (no LD):** $R \approx I$. All eigenvalues ≈1, so $\mathrm{ER}(R)\approx p$ (maximal). This suggests many degrees of freedom, but here each SNP contributes independently. In practice, a single strong causal SNP would stand out and yield a clear posterior mode (others have low posterior inclusion). However, if **multiple signals** are present of similar size, an independent-feature scenario means any permutation of those signals is equally likely *a priori*. The posterior could in theory factorize across signals (one mode per combination), but because effects add independently, the landscape is actually **unimodal per signal** – IBSS can identify each effect separately. Thus, *independence gives high ER but low "basin" complexity* (the global optimum is unique, albeit there is a combinatorial number of trivial variants of the same optimum if signals are exchangeable). Our metric remains high, reflecting many *potential* configurations; the crucial difference is that truly independent signals don't produce *distinct* peaks – they sum into one broad peak in the joint space. *(b)* **One large LD block:** All SNPs in strong LD (R nearly rank-1). Here $\mathrm{ER}(R)\approx1$. The posterior effectively has **one basin** – there is essentially one "effective variable" (the haplotype block) driving the signal. However, identifying the causal SNP is extremely difficult because many SNPs are perfectly correlated. This scenario yields *one broad peak* (the next metric will capture its breadth). Between these, more intermediate LD (many partially overlapping clusters) yields $1 < \mathrm{ER}(R) < p$, indicating

multiple basins. For instance, if R has 3–4 semi-independent LD blocks or networks, $\mathrm{ER}(R) \approx3\text{-}4$ – the posterior might have modes corresponding to a causal variant in each block.

- **Strengths & limitations:** The spectral entropy (effective rank) is smooth and mathematically well-founded. It naturally handles permutation and continuous changes in LD. It also connects to known difficulty measures: a design with *high condition number* (nearly singular R) will have an eigenvalue near 0, lowering $\mathrm{ER}(R)$ and warning of multiple equivalent solutions. One limitation is that $\mathrm{ER}(R)$ weights all eigen-directions equally; a scenario with one huge LD block and a few tiny independent clusters might yield an intermediate value (since the large block's eigenvalue dominates, ER may be closer to 1 than the raw count of clusters). In such a case, our metric would suggest fewer basins – arguably correct, since one large correlated block might monopolize the posterior, and the tiny clusters contribute minor alternative modes. Still, if one cares about *any* distinct mode (even a minor one), an alternate definition could, for example, count significant eigenvalues above a threshold. We avoid thresholds for continuity, but note that $\mathrm{ER}(R)$ can be supplemented with a *spectrum plot* to reveal the eigenvalue drop-off. Another edge case: purely independent SNPs (maximal ER) technically allow many possible causal configurations, but as noted, the posterior isn't necessarily multimodal in practice because without LD, one configuration will clearly maximize likelihood. Thus $\mathrm{ER}(R)$ may **overestimate** basins in trivial independent cases. However, in practice one often has at least some LD; truly independent signals simply yield one mode per causal SNP. Overall, LD eigenentropy is a useful proxy for how *complex* the posterior landscape might be – **low values signal a "rugged" landscape with few deep basins (often one, albeit wide), whereas intermediate values signal multiple comparable peaks.**

Alternative formulations: We chose a spectral measure for elegance, but other continuous LD-based measures can capture between-basin complexity. For example, one could use the **submodularity ratio** from feature selection theory [4] [5]: $\gamma_{k} = \min_{|S|\le k} \frac{\sum_{j\in S^c} f({j}\mid S)}{f(S\cup S^c) - f(S)}$ (for an appropriate gain function f, like variance explained). γ_k quantifies how close f is to submodular (greedy-optimal) – if γ_k is small, the objective has valleys and multiple optima. In fact, γ_k is bounded below by the smallest $2k$-sparse eigenvalue of X^TX [5], relating again to R's spectrum. While computing γ_k explicitly is hard, conceptually a **low submodularity ratio** corresponds to high between-basin complexity (many rival solutions). Another idea is to model LD as a graph and measure its community structure. Highly **modular LD graphs** (distinct communities of SNPs with high within, low between LD) imply multiple well-separated basins – one per community. A continuous metric here could be the **algebraic connectivity** of the LD graph's Laplacian (second-smallest eigenvalue): if the graph nearly disconnects into clusters, this eigenvalue $\lambda_2 \approx 0$ (indicating multiple basins). However, graph methods typically involve choosing an LD threshold to define edges, which breaks continuity and permutation-invariance unless done carefully (one could use weighted graph partitions instead). For simplicity and generality, we stick with the eigen-entropy $\mathrm{ER}(R)$ as our primary metric for between-basin complexity.

2. Within-Basin Complexity: Posterior Ambiguity in a Single Mode

Definition: *Within-basin complexity* gauges how **wide and ambiguous** the posterior is *inside* one prominent basin. Even if the global posterior has a single mode (one "basin of attraction"), that mode might not be sharp – it could encompass many nearly interchangeable variants. In fine-mapping terms, this corresponds

to the **credible set size**: how many SNPs have high posterior inclusion probability (PIP) because they are almost indistinguishable given the data [6]. A high within-basin complexity means that even after focusing on the correct region, identifying the true causal SNP is difficult – the model can **shift around within the basin** without much loss, due to correlations. We require a metric that reflects this *identifiability* issue using only X or R (no outcome), is continuous, and not tied to algorithm convergence (purely a property of LD).

Candidate Metric – LD Ambiguity Score (Average LD Degree): We propose to quantify how strongly each SNP is **redundant with others** on average. Specifically, for each SNP *j* define an *LD score* (as in LD Score regression literature) excluding itself:

$$L_j \;=\; \sum_{i \neq j} r_{ij}^2,$$

the sum of squared correlations of SNP *j* with all other SNPs. L_j measures how "entangled" SNP *j* is in the correlation web – a large L_j means *j* has many close proxies (high r^2 neighbors). Now consider a summary statistic of the L_j distribution, for instance the **mean** or **95th percentile** of L_j. A simple choice is the *mean LD score*:

$$C_{\text{within}} \;=\; \frac{1}{p}\sum_{j=1}^{p} L_j \;=\; \frac{1}{p}\sum_{j \neq i} r_{ij}^2,$$

which is just the average pairwise r^2 among SNPs (since each pair (i,j) appears in two L sums, C_{within} equals the average off-diagonal r^2). C_{within} close to 0 means most SNPs are largely independent (within any single "basin" a causal SNP stands alone, clearly identifiable). A higher C_{within} means on average SNPs have strong correlations with others, indicating that if one of them were causal, there would be several nearly equivalent proxies – **wide credible sets**. This metric is permutation-invariant and continuous in R's entries. It is also agnostic to any particular algorithm; it purely reflects LD.

- **Rationale:** If a causal signal falls in a cluster of tightly correlated SNPs, the best one can say is *"it's probably one of these 10 SNPs"*. The posterior within that basin will assign similar high probabilities to all 10, yielding a large credible set (high ambiguity) [6]. By contrast, if a causal SNP has low correlation with others, the posterior in that basin will sharply peak at that SNP (small credible set). Our proposed C_{within} captures the *general propensity* for such ambiguity by averaging over all SNPs. It effectively asks: *How interconnected is the LD network?* A highly connected (high average r^2) network means each SNP has look-alikes – any true signal would be mirrored by others (posterior spread out). A sparse network means each signal is more unique (posterior concentrated). Notably, this is related to the concept of "LD *width*": long-range or extensive LD causes ambiguity. Indeed, in practice credible sets are *"uncertainty sets of plausible causal variants within a highly correlated region"* [7] [6]. Our metric puts a number on "how highly correlated" the region is on average.

- **Properties:** C_{within} increases with both the **size and density of LD clusters**. For an extreme case of one perfect LD block of size m, each SNP has $r_{ij}=1$ with $m-1$ others, so $L_j = m-1$ for all j. The average off-diagonal r^2 is $= (m-1)/ (p-1)$ if the whole region is one block. As m grows, C_{within} approaches 1. Thus a single huge block yields $C_{\text{within}} \approx1$, reflecting that within that basin, the posterior is maximally spread (all SNPs are equally

plausible). If instead there are many small independent blocks, within each block correlations might still be high, but since the average is taken over all pairs, C_{within} will be lower. For example, consider 10 blocks of 5 SNPs each, with perfect LD within each block but none between. Within each block $L_j=4$, but across the whole region many pairs are uncorrelated. The average r^2 would be $(10 * \text{(within-block pairs)} * 1 + \text{(between-block pairs)} * 0) / \binom{50}{2}$. Numerically this is much lower than the one-block case. Thus C_{within} is sensitive to the overall prevalence of high LD pairs, not just the maximum cluster. In many scenarios, that is desirable – it gives a sense of how *pervasive* the ambiguity is in the region. However, it might **underestimate** the within-basin uncertainty if just one part of the region has a large LD cluster and the rest is sparse. In such a case, one might prefer a *maximal LD score* metric: $C'{\text{within}} = \max_j L_j$ *(the worst-case cluster size in squared-correlation units). $\max_j L_j$ directly signals the largest credible set one might face (since the SNP with highest L_j has the most "twins"). This metric is also continuous (a smooth function of R entries) and permutation-invariant. It will equal $m-1$ for a block of size m, or more generally reflect the strongest local LD concentration. The* ***trade-off*** *is that $\max L_j$ focuses on the single most ambiguous basin, whereas the* ***mean $\sum r_{ij}^2$*** *spreads the influence of all correlations. Depending on context, either could be reported. We emphasize the average C* here for its connection to well-known concepts (it's essentially the average r^2 across the region).}

- **Connections to theory:** This within-basin complexity relates to the **"signal identifiability"** in sparse recovery. In classical terms, if two predictors are highly correlated, a sparse regression problem has an *identifiability issue*: the solution (or posterior) cannot distinguish which is non-zero (both yield similar fit). Metrics like mutual coherence (maximum $|r_{ij}|$) or the L_2 norm of correlations have been used to bound identifiability. For instance, the **mutual coherence** $\mu = \max_{i\neq j}|r_{ij}|$ is a simple continuous measure of LD strength; if μ is close to 1, any causal signal will have nearly indistinguishable alternatives, blowing up credible sets. Our C_{within} can be seen as a *squared-average coherence*. Another related concept is the **Restricted Isometry Property (RIP)** constant: the largest δ such that any subset of columns up to size L approximately preserves length (no exact multicollinearity). If LD is strong, the RIP constant is high (close to 1), meaning a subset of SNPs can mimic each other's effects – again reflecting large within-basin ambiguity. While computing δ requires threshold-like worst-case analysis, our average r^2 is a softer proxy. It connects to the notion of **posterior credible set size**: approaches like Hutchinson et al. (2020) explicitly simulate credible sets based on LD structure [8] [6]. They find that credible set coverage is largely determined by the correlation structure – *"credible sets capture the uncertainty of finding the true causal variant within a highly correlated region"* [7]. In essence, C_{within} measures how *highly correlated on average* the region is, hence how large those uncertainty sets might be.

- **Strengths & limitations:** The LD ambiguity score is straightforward to compute and interpret. It requires no modeling assumptions beyond the genotype correlation and is fully agnostic to any phenotype. It's also **continuous** – small changes in correlations gradually change the score, avoiding discontinuities that a hard threshold (say, counting how many $r > 0.8$ neighbors each SNP has) would introduce. One limitation is that it does **not differentiate nearby vs distant correlation** – all pairs contribute equally to r^2 sums. In genomic contexts, LD is often local (decays with distance), so one might refine the metric by considering only a window or distance-decay weighting. But that re-introduces thresholds or tunings. As defined, C_{within} treats a spread-out moderate LD (e.g. many pairs with $r^2=0.3$) the same as a few tight pairs with $r^2=0.9$ in terms of average. In practice, moderate but widespread LD can also create ambiguity (the posterior might smear out over

a region if no single SNP stands out strongly), so including all r^2 is reasonable. If needed, one can complement C_{within} with something like $\max r_{ij}^2$ (to know the strongest LD) to get a fuller picture. Finally, note that high C_{within} does not *guarantee* a large credible set in a given analysis – if the true effect is very strong at one SNP, it might still dominate the others – but it flags the *intrinsic difficulty*: **a high average correlation environment is fertile ground for posterior ambiguity** if effect sizes are comparable.

Comparing the Two Metrics and Interpretation

These two metrics capture **different aspects of complexity** in the SuSiE/IBSS posterior landscape:

- **Between-basin complexity** (e.g. effective LD rank $\mathrm{ER}(R)$) ~ **how many distinct modes** might the posterior have? It is high when the LD structure suggests multiple separated "signal locations" or combinations could explain the data (multimodal posterior), and low when only one configuration (or a couple) is plausible. It relates to **global multimodality** and the possibility of **multiple local optima** for the IBSS algorithm's objective. A very low value (near 1) implies a *unimodal* or near-unimodal posterior (but possibly one very broad mode), whereas a high value indicates potential **posterior multi-modality** (several competitive configurations in different LD blocks or dimensions). This connects to whether the feature-selection objective is close to submodular (yielding a clear optimum) or highly non-submodular (many optima) [4] [5].

- **Within-basin complexity** (C_{within} or related measures) ~ **how diffuse is each mode**? Even if there's a single cluster of signals, if that cluster contains many correlated SNPs, the posterior mass will be spread among them (one broad peak). This reflects **local uncertainty/identifiability**. It correlates with credible set size: high within-basin complexity means large credible sets (many near-equivalent causal candidates) [6], whereas low complexity means pinpointing a causal SNP is easier (small credible sets). Technically, this is related to the **local curvature** of the posterior around the mode – a flatter mode (due to correlated parameters) means more ambiguity.

It's worth noting the **interplay** of these metrics. They are not entirely independent: an LD structure with one gigantic block will yield low between-basin (≈ 1) but high within-basin (very large C_{within}). A structure with many tiny independent blocks yields high between-basin (many potential modes) but each mode is sharp (low within-basin, since each cause is isolated). The "hardest" scenarios for inference often lie in between – **moderate between-basin and high within-basin** simultaneously: e.g. several LD blocks, each of which still contains 10–20 highly correlated SNPs. Then there are multiple modes *and* each mode is internally ambiguous. IBSS could oscillate between picking signals in different blocks (multiple local optima), and even within one block it might not confidently choose the right SNP (flat likelihood surface). Such situations indeed occur with "blurry" LD – e.g. haplotype blocks that overlap or regions of high but not complete LD that extend over large genomic distances. Our metrics would flag this: $\mathrm{ER}(R)$ might be, say, 3 or 4 (several directions), and C_{within} might be, say, 0.5 or 0.6 (on average, 50–60% r^2 among SNPs – quite strong). In contrast, a simpler scenario (one clear signal in one tight cluster) might have $\mathrm{ER}(R)\approx 1.2$ (essentially one mode) but $C_{\text{within}}\approx 0.8$ (the cluster is tight, giving a wide credible set but no alternative modes). Both metrics together thus give a more complete picture of difficulty.

Existing theory ties: In Bayesian variational inference, a unimodal variational family (like a factorized Gaussian in SuSiE's variational EB step) may struggle with multimodal posteriors – effectively it might lock

onto one mode and miss others [9] [10]. A high between-basin metric warns that the true posterior is far from unimodal, so the IBSS's mean-field approximation might yield one posterior mode's credible set and neglect another mode entirely. This can lead to **false confidence** (e.g., IBSS might report a high PIP cluster while another equally plausible cluster is ignored due to the algorithm's initialization or greedy path [1]). Meanwhile, high within-basin complexity indicates that even if IBSS finds the right general area, the variational posterior will be very flat/spread – the ELBO surface has a broad plateau. Convergence might slow, and the result might be very sensitive to prior or hyperparameters (since many SNPs share the likelihood, small prior tweaks could shift allocation of probability among them).

From a **sparse recovery** perspective, high between-basin complexity often means the design matrix fails certain uniqueness conditions (like the restricted eigenvalue condition [5]) – multiple sparse solutions exist with similar fit. High within-basin complexity corresponds to having predictors with high mutual coherence – any sparse solution is non-unique unless extra constraints pick one. These metrics, therefore, recast classical design difficulty measures into a form relevant for Bayesian posterior landscapes.

Implications for IBSS and Advanced Strategies

Predicting problem difficulty with these metrics not only is useful for diagnosing fine-mapping challenges, but also **suggests improvements** to inference algorithms:

- **Multi-basin exploration:** If the between-basin complexity is high (suggesting multiple posterior modes), one may need to run IBSS (or any greedy/coordinate algorithm) from multiple starting points or incorporate **multi-modal averaging**. For example, in Bayesian deep learning it's been shown that averaging solutions from different basins ("deep ensembles") approximates the true multimodal posterior better than a single-mode approximation. In the fine-mapping context, one could run SuSiE/IBSS multiple times with random initializations or perturbations to discover alternative high-ELBO solutions (different credible sets) and then **average across basins**. Recent approaches like *MultiSWAG* (multi-sample weight-averaged Gaussian) explicitly marginalize over multiple modes and have proven effective in reducing generalization error [11]. An analogue for fine-mapping might keep multiple candidate sets of effects (basins) and form a Bayesian model average to compute PIPs that reflect uncertainty across modes. The between-basin metric could even guide a **temperature annealing** schedule – e.g., a high complexity might warrant starting with a "heated" posterior (flattened likelihood) to allow jumps between basins, gradually cooling to focus on the best mode.

- **Mirror descent and convexification:** Mirror descent algorithms can sometimes navigate complicated posteriors by working in a transformed (dual or density) space. In fact, *Particle Mirror Descent (PMD)* is a variational inference method that maintains a set of particles and updates them via mirror descent in the space of probability distributions [12] [10]. This approach can capture multi-modality because it does not force a single parametric form – the particles can spread over different modes [10]. If $\mathrm{ER}(R)$ is high, a mirror-descent-based IBSS might, for instance, treat the variational posterior as a mixture and update mixture weights and components, rather than a single Gaussian. Similarly, one could incorporate a **submodular relaxation**: for example, using a convex surrogate objective that is tighter when the submodularity ratio is low. One idea is to add an **entropy or divergence term** to the IBSS updates to discourage premature commitment to one mode. This is akin to simulated annealing or deterministic annealing in clustering, where a high entropy (or temperature) keeps the posterior landscape smooth enough to traverse between basins.

- **Temperature annealing:** As noted, adjusting the Bayesian posterior temperature T (raising or lowering the influence of the likelihood vs prior) can smooth the landscape [13] [14]. A *"warm" posterior* (high T) essentially down-weights the data relative to the prior, making the posterior more diffuse and unimodal, which can help algorithms explore. Then one can gradually cool to $T=1$. If we know a priori that LD structure is complex (high metrics), we might start IBSS on a tempered posterior to avoid getting stuck. Conversely, **cold posteriors** (low T) concentrate on high-likelihood modes and might be useful to refine within a basin once identified [15]. The idea of **continuation methods** (solve an easier, smoothed problem first) is well-established in non-convex optimization, and here the metrics could trigger such strategies.

- **Mirror descent for sparsity:** Another angle is using mirror descent on the *feature inclusion probabilities* (the gamma parameters in SuSiE). IBSS essentially does coordinate ascent on each single-effect's inclusion and effect distribution. In a difficult LD scenario, a naive coordinate update might ping-pong between correlated SNPs. A mirror-descent update (treating it as an optimization on the probability simplex of PIPs, using an entropy mirror map) could naturally enforce a form of diminishing returns. In fact, a fully Bayesian view sees the posterior as the solution of a **convex problem in density space** [16] [17]; mirror descent provides a way to optimize that while maintaining entropy to capture uncertainty. While these ideas are speculative, the connection is that our metrics hint at *why* IBSS might fail (multiple comparable optima or flat regions) and thus what interventions might help (exploration vs exploitation phases).

In summary, by quantifying **between-basin** and **within-basin complexity**, we gain insight into the *posterior landscape* of a fine-mapping problem before even seeing the phenotype data. A high between-basin metric warns of possible **multiple modes** (the method might need to search more broadly or report multiple credible sets), while a high within-basin metric warns of **low identifiability** (expect large credible sets and uncertainty even within one mode) [6]. These metrics are continuous, permutation-invariant, and grounded in the geometry of the genotype correlation matrix. They connect to theoretical concepts like submodularity (greedy optimality relates to having essentially one basin [4] [5]) and to practical issues like variational approximation quality (which suffers when the true posterior is multimodal or flat [9] [10]). Ultimately, such metrics could not only flag difficult regions for researchers (prompting more careful analysis or inclusion of external information to resolve ambiguity) but also inspire algorithmic enhancements – from multi-start strategies and posterior tempering to entirely new frameworks that better handle **multi-basin posteriors** [14] and **wide basins** in high-LD regions.

Sources:

- Wang et al. (2020) – SuSiE fine-mapping and IBSS algorithm introduction [18] [1].
- Das & Kempe (2018) – Submodularity ratio and sparse eigenvalue connection [5].
- Hutchinson et al. (2020) – credible sets and LD uncertainty [6].
- Abhimanyu Das, David Kempe. *Approximate Submodularity for subset selection* – greedy performance vs eigenvalues [5].
- Particle Mirror Descent (Dai et al. 2016) – capturing multimodal posteriors with mirror descent [10].
- Maddox et al. (2019) – MultiSWAG and multi-basin Bayesian model averaging in deep learning.
- Bayesian posterior tempering and multimodality discussion [14].

---

1 Fine-mapping from summary data with the “Sum of Single Effects ...
https://pmc.ncbi.nlm.nih.gov/articles/PMC9337707/

2 18 Fine-mapping from summary data with the “Sum of Single Effects” model | PLOS Genetics
https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010299

3 [PDF] Matrix Information Theory for Self-Supervised Learning - arXiv
https://arxiv.org/pdf/2305.17326

4 5 jmlr.org
https://jmlr.org/papers/volume19/16-534/16-534.pdf

6 Improving the coverage of credible sets in Bayesian genetic fine-mapping | PLOS Computational Biology
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007829

7 BEATRICE: Bayesian fine-mapping from summary data using deep ...
https://academic.oup.com/bioinformatics/article/40/10/btae590/7808857

8 Improving the coverage of credible sets in Bayesian genetic fine ...
https://www.researchgate.net/publication/340612482_Improving_the_coverage_of_credible_sets_in_Bayesian_genetic_fine-mapping

9 10 12 16 17 proceedings.mlr.press
http://proceedings.mlr.press/v51/dai16.pdf

11 13 14 15 Bayesian Deep Learning - Abridged.pdf
file://file_000000001d3c71f8a9e8866af05e7f5b