Beyond Perturbation Stability: LP Recovery Guarantees for MAP Inference on Noisy Stable Instances

Hunter Lang[*]
MIT
hjl@mit.edu

Aravind Reddy[*]
Northwestern University
arareddy@u.northwestern.edu

David Sontag
MIT
dsontag@mit.edu

Aravindan Vijayaraghavan
Northwestern University
aravindv@northwestern.edu

Abstract

Several works have shown that *perturbation stable* instances of the MAP inference problem in Potts models can be solved exactly using a natural linear programming (LP) relaxation. However, most of these works give few (or no) guarantees for the LP solutions on instances that do not satisfy the relatively strict perturbation stability definitions. In this work, we go beyond these stability results by showing that the LP approximately recovers the MAP solution of a stable instance even after the instance is corrupted by noise. This "noisy stable" model realistically fits with practical MAP inference problems: we design an algorithm for finding "close" stable instances, and show that several real-world instances from computer vision have nearby instances that are perturbation stable. These results suggest a new theoretical explanation for the excellent performance of this LP relaxation in practice.

1 Introduction

In this work, we study the MAP inference problem in the *ferromagnetic Potts model*, which is also known as uniform metric labeling (Kleinberg & Tardos, 2002). Given a graph $G = (V, E)$, this problem is:

$$\underset{x:V \to [k]}{\text{minimize}} \sum_{u \in V} c(u, x(u)) + \sum_{(u,v) \in E} w(u,v) \mathbb{1}[x(u) \neq x(v)].$$

Here we are optimizing over *labelings* $x : V \to [k]$ where $[k] = \{1, 2, \ldots, k\}$. The objective is comprised of "node costs" $c : V \times [k] \to \mathbb{R}$, and "edge weights" $w : E \to \mathbb{R}_{>0}$; a labeling x pays the cost $c(u, i)$ when it labels node u with label i and pays $w(u, v)$ on edge

(u, v) when it labels u and v differently. This problem is NP-hard for variable $k \geq 3$ (Kleinberg & Tardos, 2002) even when the graph G is planar (Dahlhaus et al., 1992). However, there are several efficient and empirically successful approximation algorithms for the MAP inference problem—such as TRW (Wainwright et al., 2005) and MPLP (Globerson & Jaakkola, 2008)—that are related in some way to the *local LP relaxation*, which is also sometimes called the *pairwise LP* (Wainwright & Jordan, 2008; Chekuri et al., 2001). This LP relaxation returns an approximate MAP solution for most problem instances. However, when the parameters of these models are learned so as to enable good structured prediction, often the LP relaxation exactly or almost exactly recovers the MAP solution (Meshi et al., 2019). The connection between the LP relaxation and commonly used approximate MAP inference algorithms then leads to the following compelling question, which is of great practical relevance for understanding the "tightness" of the LP solution (informally, how close the LP solution is to the MAP solution).

Can we explain the exceptional performance of the local LP relaxation in recovering the MAP solution in practice?

Several works have studied different conditions that imply the local relaxation or related relaxations are tight (e.g., Kolmogorov & Wainwright, 2005; Wainwright & Jordan, 2008; Thapper et al., 2012; Weller et al., 2016; Rowland et al., 2017). Recent work on tightness of the local relaxation has focused on a class of several related conditions known as *perturbation stability*. Intuitively, an instance is perturbation stable if the solution x^* to the MAP inference problem is unique, and moreover, x^* is the unique solution even when the edge weights w are multiplicatively perturbed by a certain adversarial amount (Bilu & Linial, 2010). This structural assumption about the instance (G, c, w) captures the intuition that, on "real-world" instances, the ground-truth solution is stable and does not change much when the weights are slightly perturbed.

For constants $\beta, \gamma \geq 1$, we say that w' is a (β, γ)-
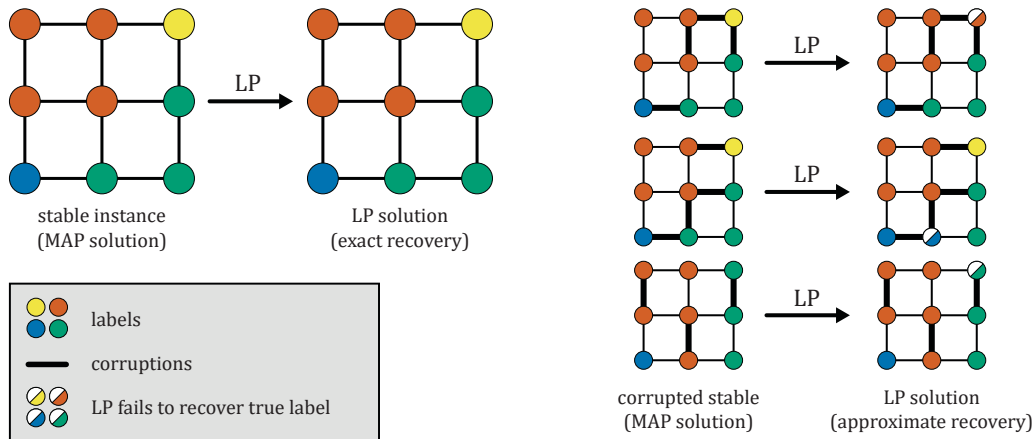
---

[*]Denotes equal contribution.

Figure 1: Left: prior work (Lang et al., 2018) showed that a *stable instance* can be exactly solved efficiently. Colors indicate the label of each vertex in the MAP solution x^*. On stable instances, solving the LP relaxation (represented by the arrow) recovers the MAP solution. However, real-world instances are *not* suitably stable for this result to apply in practice (Lang et al., 2019). Right: in this work, we show that solving the LP relaxation on a (slightly) *corrupted* stable instance (corruptions shown as bold edges) *approximately* recovers the original MAP solution. This is true even if the corruption changes the MAP solution (as in the bottom example). In other words, we prove that "easy" instances are still approximately easy even after small perturbations.

perturbation of the weights w if $\frac{1}{\beta} \cdot w(u, v) \leq w'(u, v) \leq \gamma \cdot w(u, v)$ for all $(u, v) \in E$. Suppose x^* is the unique MAP solution to the instance (G, c, w). Then, we say (G, c, w) is a (β, γ)-stable instance if x^* is also the unique MAP solution to every instance (G, c, w') where w' is a (β, γ)-perturbation of w. Lang et al. (2018) showed that when (G, c, w) is $(2, 1)$-stable, the solution to the local LP relaxation is *persistent* i.e., the LP solution exactly recovers the MAP solution x^*.

While theoretically interesting, $(2, 1)$-stability is a strict condition that is unlikely to be satisfied in practice: the solution x^* is not allowed to change *at all* when the weights are perturbed. No real-world instances have yet been shown to be $(2, 1)$-stable (Lang et al., 2019). Moreover, the LP relaxation is also not persistent on most of those instances. However, the solution of the local LP relaxation is still *nearly persistent* i.e., the LP solution is very close to the MAP solution x^* (see Definition 3.1 for a formal definition). Those examples made it clear that theory must go beyond perturbation stability to explain this phenomenon of near-persistence that is prevalent in practice (see e.g., Sontag, 2010; Shekhovtsov et al., 2017).

Why is the LP relaxation nearly persistent on MAP inference instances in practice?

There are several theoretical frameworks to explain exactness or tightness of LP relaxations, such as total unimodularity, submodularity (Kolmogorov & Wainwright, 2005), and perturbation stability (Lang et al., 2018, 2019), as well as structural assumptions about

the graph G (Wainwright & Jordan, 2008), or combined assumptions about G and the form of the objective function (Weller et al., 2016; Rowland et al., 2017). However, these frameworks can not be used to prove near-persistence.

Figure 1 (informally) shows our main result. The left side depicts the previous result of Lang et al. (2018): if the instance is $(2, 1)$-stable (a fairly strong structural assumption), the LP relaxation exactly recovers the full solution x^*. This result is limited because real-world instances have been shown to not satisfy $(2, 1)$-stability (Lang et al., 2019). The right side shows our main result: if the instance is a *slightly corrupted* $(2, 1)$-stable instance, the LP relaxation still *approximately* recovers the solution x^* to the stable instance.

Intuitively, we may expect a real-world instance to be "close" to a stable instance (i.e., to be a "corrupted stable" instance, as in Figure 1) even if the instance itself is not stable. We design an algorithm to check whether this is the case. We find that on several real examples, sparse and small-norm changes to the instance make it appropriately stable for our theorems to apply. In other words, we certify that these real instances are close to stable instances. For these instances, our theoretical results explain why the LP relaxation approximately recovers the MAP solution.

More formally, we assume that there is some *latent* stable instance $(G, \bar{c}, \bar{w})$, and that the observed instance $(G, \hat{c}, \hat{w})$ is a noisy version of $(G, \bar{c}, \bar{w})$ that is close to it. Let $\hat{x}$ be the solution to the local LP on the observed

instance $(G, \hat{c}, \hat{w})$, and let $\bar{x}$ be the (unknown) MAP solution on the unseen stable instance $(G, \bar{c}, \bar{w})$. We prove that under certain conditions, the LP solution $\hat{x}$ is *nearly persistent* i.e., the Hamming error $\|\hat{x} - \bar{x}\|_1$ is small (see Definition 3.1). In other words, the local LP solution to the observed instance approximately recovers the latent integral solution $\bar{x}$.

We complement this by studying a natural generative model that generates noisy stable instances which, with high probability, satisfy the above conditions for *near persistency*. In other words, the observed instance $(G, \hat{c}, \hat{w})$ is obtained by random perturbations to the latent stable instance $(G, \bar{c}, \bar{w})$, and the LP relaxation approximately recovers the MAP solution to the latent instance with high probability.

Our theoretical results imply that the local LP approximately recovers the MAP solution when the observed instance is close to a stable instance. Our empirical results suggest that real-world instances are very close to stable instances. These results together suggest a new explanation for the near-persistency of the solution of the local LP relaxation for MAP inference in practice. To prove these results and derive our algorithm for finding a "close-by" stable instance, we make several novel technical contributions, which we outline below.

Technical contributions.

- In Section 4, we generalize the $(2, 1)$-stability result of Lang et al. (2018) to work under a much weaker assumption, which we call $(2, 1)$-expansion stability. That is, we prove the local LP is tight on $(2, 1)$-expansion stable instances. Additionally, given the instance's MAP solution, $(2, 1)$-expansion stability is efficiently checkable. To the best of our knowledge, most other perturbation stability assumptions are not known to satisfy this desirable property. This generalization is crucial for the efficiency of our algorithm for finding stable instances that are close to a given observed instance.

- In Section 5, we give a simple extension of $(2, 1)$-expansion stability called $(2, 1, \psi)$-expansion stability. We prove it implies a "curvature" result around the MAP solution $\bar{x}$. On instances that satisfy this condition, if a labeling $\hat{x}$ is close in objective value to $\bar{x}$, it must also be close in the solution space. This result lets us translate between objective gap and Hamming distance.

- In Section 6, we study a natural generative model where the observed instance is generated from an arbitrary *latent stable* instance by random (sub-Gaussian) perturbations to the costs and weights. We prove that, with high probability, every feasible LP solution takes close objective values on the

latent and observed instances. The proof uses a rounding algorithm for metric labeling in a novel way to obtain stronger guarantees. When combined with our other results, this proves that when the latent instance is $(2, 1, \psi)$-expansion stable, the LP solution is nearly persistent on the observed instance with high probability. These results suggest a theoretical explanation for the phenomenon of near-persistence of the LP solution in practice.

- We design an efficient algorithm for finding $(2, 1, \psi)$-expansion stable instances that are "close" to a given instance $(G, \hat{c}, \hat{w})$ in Section 7. To the best of our knowledge, this is the first algorithm for finding close-by stable instances, and is also an efficient algorithm for checking $(2, 1, \psi)$-expansion stability. This algorithm allows us to check whether real-world instances can plausibly be considered "corrupted stable" instances as shown in Figure 1.

- We run our algorithm on several real-world instances of MAP inference in Section 8, and find that the observed instances $(G, \hat{c}, \hat{w})$ often admit close-by $(2, 1, \psi)$-stable instances $(G, \bar{c}, \bar{w})$. Moreover, we find that the local LP solution $\hat{x}$ typically has very close objective to $\bar{x}$ in $(G, \bar{c}, \bar{w})$. Our curvature result for $(2, 1, \psi)$-stable instances thus gives an explanation for the tightness of the local LP relaxation on $(G, \hat{c}, \hat{w})$.

2 Related work

Perturbation stability. Several works have given recovery guarantees for the local LP relaxation on perturbation stable instances of uniform metric labeling (Lang et al., 2018, 2019) and for similar problems (Makarychev et al., 2014; Angelidakis et al., 2017).

Lang et al. (2019) give partial recovery guarantees for the local LP when parts (*blocks*) of the observed instance satisfy a stability-like condition, and they showed that practical instances have blocks that satisfy their condition. However, the required *block stability* condition in turn depends on certain quantities related to the LP dual. This is unsatisfactory since this does not explain when and why such instances are likely to arise in practice. For a more extensive treatment of the subject, we refer the reader to the "Perturbation Resilience" chapter from Roughgarden (2021).

Easy instances corrupted with noise. Our random noise model is similar to several planted average-case models like stochastic block models (SBMs) considered in the context of problems like community detection, correlation clustering and partitioning (see e.g., McSherry, 2001; Abbe, 2018; Globerson et al., 2015).

Instances generated from these models can also be seen as the result of random noise injected into an instance with a nice block structure that is easy to solve. Several works give exact recovery and approximate recovery guarantees for semidefinite programming (SDP) relaxations for such models in different parameter regimes (Abbe, 2018; Guédon & Vershynin, 2016). In our model however, we start with an *arbitrary* stable instance as opposed to an instance with a block structure (which is trivial to solve). Moreover, we are unaware of such analysis in the context of linear programs. Please see Section 6 for a more detailed comparison. To the best of our knowledge, we are the first to study instances generated from random perturbations to stable instances.

Partial optimality algorithms. Several works have developed fast algorithms for identifying parts of the MAP assignment. These algorithms output an approximate solution $\hat{x}$ and a set of vertices where $\hat{x}$ provably agrees with the MAP solution x^* (e.g., Kovtun, 2003; Shekhovtsov, 2013; Swoboda et al., 2016; Shekhovtsov et al., 2017). Like these works, our results also prove that an approximate solution $\hat{x}$ has small error $|\hat{x} - x^*|$. However, these previous approaches are more concerned with designing fast algorithms for finding such $\hat{x}$. In contrast, we focus on giving structural conditions that explain why a *particular* $\hat{x}$ (the solution to the local LP relaxation) often approximately recovers x^*. Our algorithm in Section 7 is thus not meant as an efficient method for certifying that $|\hat{x} - x^*|$ is small, but rather as a method for checking whether our structural condition (that the observed instance is close to a stable instance) is satisfied in practice.

3 Preliminaries

In this section we introduce our notation, define the local LP relaxation for MAP inference, and give more details on perturbation stability. As in the previous section, the MAP inference problem in the ferromagnetic Potts model on the instance (G, c, w) can be written in *energy minimization* form as:

$$\underset{x:V\to[k]}{\text{minimize}} \sum_{u\in V} c(u, x(u)) + \sum_{(u,v)\in E} w(u,v)\mathbb{1}[x(u)\neq x(v)].$$
$$(1)$$

Here x is an *assignment* (or labeling) of vertices to labels i.e. $x : V \to \{1, 2, \ldots, k\}$. We can identify each labeling x with a point $(x_u : u \in V; x_{uv} : (u,v) \in E)$, where each $x_u \in \{0,1\}^k$ and each $x_{uv} \in \{0,1\}^{k\times k}$.

In this work, we consider all node costs $c(u,i) \in \mathbb{R}$ and all edge weights $w(u,v) > 0$. We note that this is equivalent to the formulation where all node costs and edge weights are non-negative (Kleinberg & Tardos,

2002). See Appendix A for a proof of this equivalence.

We encode the node costs and the edge weights in a vector $\theta \in \mathbb{R}^{nk+mk^2}$ where $n = |V|$ and $m = |E|$ s.t. $\theta(u, i) = c(u, i), \theta(u, v, i, j) = w(u, v)\mathbb{1}[i \neq j]$. Then the objective can be written as $\langle \theta, x\rangle$. We set $x_u(i) = 1$ when $x(u) = i$, and 0 otherwise. Similarly, we set $x_{uv}(i, j) = 1$ when $x(u) = i$ and $x(v) = j$, and 0 otherwise. Where convenient, we use x to refer to this point rather than the labeling $x : V \to [k]$. We can then rewrite (1) as:

$$\min_x. \sum_{u\in V}\sum_{i=1}^{k} c(u,i)x_u(i) + \sum_{(u,v)\in E} w(u,v)\sum_{i\neq j} x_{uv}(i,j)$$

$$\text{subject to: } \sum_{i=1}^{k} x_u(i) = 1 \qquad \forall\, u \in V$$

$$\sum_{i=1}^{k} x_{uv}(i,j) = x_v(j) \quad \forall\, (u,v) \in E,\; j \in [k]$$

$$\sum_{j=1}^{k} x_{uv}(i,j) = x_u(i) \quad \forall\, (u,v) \in E,\; i \in [k]$$

$$x_{uv}(i,j) \in \{0,1\} \qquad \forall\, (u,v),\; (i,j)$$

$$x_u(i) \in \{0,1\} \qquad \forall\, u,\; i.$$

This is equivalent to (1), and is an integer linear program (ILP). By relaxing the integrality constraints from $\{0, 1\}$ to $[0, 1]$, we obtain the *local LP relaxation*:

$$\min_{x\in L(G)}. \sum_{u\in V}\sum_{i=1}^{k} c(u,i)x_u(i) + \sum_{(u,v)\in E} w(u,v)\sum_{i\neq j} x_{uv}(i,j),$$

where $L(G)$ is the polytope defined by the same constraints as above, with $x \in \{0,1\}$ replaced with $x \in [0,1]$. This is known as the *local polytope* (Wainwright & Jordan, 2008). The vertices of $L(G)$ are either *integral*, meaning all x_u and x_{uv} take values in $\{0,1\}$, or *fractional*, when some variables take values in $(0, 1)$. Integral vertices of this polytope correspond to labelings $x : V \to [k]$, so if the LP solution is obtained at an integral vertex, then it is also a MAP assignment.

If the solution x^* of this relaxation on an instance (G, c, w) is obtained at an integral vertex, we say the LP is *tight* on the instance, because the LP has exactly recovered a MAP assignment. If the LP is not tight, there may still be some vertices u where x_u^* takes integral values. In this case, if $x_u^*(i) = 1$ and $\bar{x}(u) = i$, i.e., the LP solution agrees with the MAP assignment $\bar{x}$ at vertex u, the LP is said to be *persistent* at u. $x_u^*(i) \in \{0,1\}$ does not imply the LP is persistent at u, in general. The LP solution x^* is said to be persistent if it agrees with $\bar{x}$ at every vertex $u \in V$.

Recovery error: In practice, the local LP relaxation is often not tight, but is *nearly persistent*. We will

measure the recovery error of our LP solution in terms of the "Hamming error" between the LP solution and the MAP assignment.

Definition 3.1 (Recovery error). Given an instance (G, c, w) of (1), let $\bar{x}$ be a MAP assignment, and let x^* be a solution to the local LP relaxation. The recovery error is given by (with some abuse of notation)

$$\frac{1}{2}\|x^* - \bar{x}\|_1 := \frac{1}{2}\|x_V^* - \bar{x}_V\|_1$$
$$= \frac{1}{2}\sum_{u\in V}\sum_{i\in[k]}|x_u^*(i) - \mathbb{1}[\bar{x}(u) = i]|.$$

$x_V \in \mathbb{R}^{nk}$ denotes the portion of x restricted to the vertex set V. If x^* is integral, the recovery error measures the number of vertices where x^* disagrees with $\bar{x}$. When the recovery error of x^* is 0, the solution x^* is *persistent*. We will say that the LP solution x^* is *nearly persistent* when the recovery error of solution x^* is a small fraction of n.

In our analysis, we will consider the following subset $L^*(G)$ of $L(G)$ which is easier to work with, and which contains all points we are interested in.

Definition 3.2 ($L^*(G)$). We define $L^*(G) \subseteq L(G)$ to be the set of points $x \in L(G)$ which further satisfy the constraint that $x_{uv}(i, i) = \min(x_u(i), x_v(i))$ for all $(u, v) \in E$ and $i \in [k]$.

Claim 3.3. *For a given graph G, every solution $x \in L(G)$ that minimizes $\langle \theta, x \rangle$ for some valid objective vector $\theta = (c, w)$ also belongs to $L^*(G)$. Further, all integer solutions in $L(G)$ also belong to $L^*(G)$.*

We prove this claim in Appendix A.

Our new stability result relies on the set of *expansions* of a labeling x.

Definition 3.4 (Expansion). Let $x : V \to [k]$ be a labeling of V. For any label $\alpha \in [k]$, we say that x' is an α-expansion of x if $x' \neq x$ and the following hold for all $u \in V$:

$$x(u) = \alpha \implies x'(u) = \alpha,$$
$$x'(u) \neq \alpha \implies x'(u) = x(u).$$

That is, x' may only expand the set of points labeled α, and cannot make other changes to x.

4 Expansion Stability

In this section, we generalize the stability result of Lang et al. (2018) to a much broader class of instances. This generalization allows us to efficiently check whether a real-world instance could plausibly have the structure
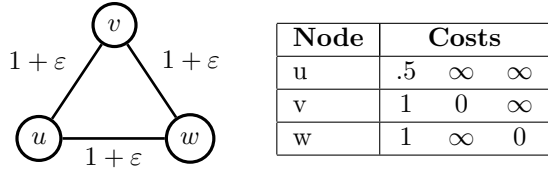

Node	Costs		
u	.5	∞	∞
v	1	0	∞
w	1	∞	0

Figure 2: $(2, 1)$-expansion stable instance that is not $(2, 1)$-stable. In the original instance (shown left), the optimal solution labels each vertex with label 1, for an objective of 2.5. This instance is not $(2, 1)$-stable: consider the $(2, 1)$-perturbation that multiples all edge weights by $1/2$. In this perturbed instance, the original solution still has objective 2.5, and the new optimal solution labels $(u, v, w) \to (1, 2, 3)$. This has a node cost of 0.5 and an edge cost of $(3 + 3\varepsilon)/2$, for a total of $2 + 3\varepsilon/2 < 2.5$. Since the original solution is not optimal in the perturbed instance, this instance is not $(2, 1)$-perturbation stable. However, note that the only expansions of the original solution (which had all label 1) that have non-infinite objective are $(u, v, w) \to (1, 2, 1)$ and $(u, v, w) \to (1, 1, 3)$. These each have objective $2.5 + \varepsilon$, which is strictly greater than the perturbed objective of the original solution. In fact, checking this single perturbation, known as the *adversarial perturbation* is enough to verify expansion stability: this instance is $(2, 1)$-expansion stable. We include the full details in Appendix B.

shown in Figure 1 (that is, whether the instance is close to a suitably stable instance).

Consider a fixed instance (G, c, w) with a unique MAP solution $\bar{x}$. Theorem 1 of Lang et al. (2018) requires that for all $\theta' \in \{(c, w') \mid w' \in \{(2, 1)\text{-perturbations of } w\}\}$, $\langle \theta', x \rangle > \langle \theta', \bar{x} \rangle$ for *all* labelings $x \neq \bar{x}$. That is, that result requires $\bar{x}$ to be the unique optimal solution in any $(2, 1)$-perturbation of the instance. By contrast, our result only requires $\bar{x}$ to have better perturbed objective than the set of *expansions* of $\bar{x}$ (c.f. Definition 3.4).

Definition 4.1 ($(2,1)$-expansion stability). Let $\bar{x}$ be the unique MAP solution for (G, c, w), and let $\mathcal{E}_{\bar{x}}$ be the set of expansions of $\bar{x}$ (see Definition 3.4). Let

$$\Theta = \{(c, w') \mid w' \in \{(2, 1)\text{-perturbations of } w\}\}$$

be the set of all objective vectors within a $(2, 1)$-perturbation of $\theta = (c, w)$. We say the instance (G, c, w) is $(2, 1)$-expansion stable if the following holds for all $\theta' \in \Theta$ and all $x \in \mathcal{E}_{\bar{x}}$:

$$\langle \theta', x \rangle > \langle \theta', \bar{x} \rangle.$$

That is, $\bar{x}$ is better than all of its expansions $x \neq \bar{x}$ in every $(2, 1)$-perturbation of the instance.

Theorem 4.2 (Local LP is tight on $(2, 1)$-expansion stable instances)**.** *Let $\bar{x}$ and $\hat{x}$ be the MAP and local LP solutions to a $(2, 1)$-expansion stable instance (G, c, w), respectively. Then $\bar{x} = \hat{x}$ i.e. the local LP is tight on (G, c, w).*

We defer the proof of this theorem to Appendix B as it is similar to the proof of Theorem 1 from Lang et al. (2018). The $(2, 1)$-expansion stability assumption is much weaker than $(2, 1)$-stability because the former only compares $\bar{x}$ to its expansions, whereas the latter compares $\bar{x}$ to *all* labelings. While the rest of our results can also be adapted to the $(2, 1)$-stability definition, this relaxed assumption gives better empirical results. Figure 2 shows an example of a $(2, 1)$-expansion stable instance that is not $(2, 1)$-stable. This shows that our new stability condition is less restrictive.

5 Curvature around MAP solution and near persistence of the LP solution

In this section, we show that a condition related to $(2, 1)$-expansion stability, called $(2, 1, \psi)$-expansion stability, implies a "curvature" result for the objective function around the MAP solution $\bar{x}$. On instances satisfying this condition, any point $\hat{x} \in L(G)$ with objective close to $\bar{x}$ also has small $\|\hat{x} - \bar{x}\|_1$, so $\hat{x}$ and $\bar{x}$ are close in solution space. In other words, if the LP solution $\hat{x}$ to a "corrupted" $(2, 1, \psi)$-expansion stable instance is near-optimal in the original $(2, 1, \psi)$-expansion stable instance (whose solution is $\bar{x}$), then the result in this section implies $\|\hat{x} - \bar{x}\|_1$ is small. This immediately gives a version of the result in the right panel of Figure 1: suppose we define an instance to be *close* to a $(2, 1, \psi)$-expansion stable $(G, \bar{c}, \bar{w})$ if its LP solution $\hat{x}$ is approximately optimal in $(G, \bar{c}, \bar{w})$. Then the curvature result implies that the LP approximately recovers the stable instance's MAP solution $\bar{x}$ for all close instances. In Section 6, we give a generative model where the generated instances are "close" according to this definition with high probability.

The $(2, 1, \psi)$-expansion stability condition, for $\psi > 0$, says that the instance is $(2, 1)$-expansion stable even if we allow all node costs $c(u, i)$ to be additively perturbed by up to ψ. This extra additive stability will allow us to prove the curvature result. This is related to the use of additive stability in Lang et al. (2019) to give persistency guarantees.

Definition 5.1 ($(2, 1, \psi)$-expansion stable)**.** For $\psi > 0$, we say an instance (G, c, w) is $(2, 1, \psi)$-expansion stable if (G, c', w) is $(2, 1)$-expansion stable for all c' with $c \leq c' \leq c + \psi \cdot \mathbf{1}$ where $\mathbf{1}$ is the all-ones vector.

The following theorem shows low recovery error i.e.,

near persistence of the LP solution on $(2, 1, \psi)$ expansion stable instances in terms of the gap in objective value.

Theorem 5.2. *Let (G, c, w) be a $(2, 1, \psi)$-expansion stable instance with MAP solution $\bar{x}$. Let $\theta = (c, w)$. Then for any $x \in L^*(G)$, the recovery error (see Def. 3.1) satisfies*

$$\frac{1}{2}\|x - \bar{x}\|_1 := \frac{1}{2}\|x_V - \bar{x}_V\|_1 \leq \frac{1}{\psi}|\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle|. \quad (2)$$

Proof (sketch). For any $x \in L^*(G)$, we construct a feasible solution $\hat{x}$ which is a strict convex combination of x and $\bar{x}$ that is very close to $\bar{x}$. Then, we apply a rounding algorithm to $\hat{x}$ to get a random integer solution h. Let $\hat{\theta}$ represent the worst $(2, 1)$-perturbation for $\bar{x}$. This is the instance where all the edges not cut by $\bar{x}$ have their weights multiplied by $1/2$. We define the objective difference using $\hat{\theta}$ as $A_h = \langle \hat{\theta}, h \rangle - \langle \hat{\theta}, \bar{x} \rangle$. First we show an upper bound for $\mathbb{E}[A_h]$ using properties of the rounding algorithm. Then we show that for any solution h in the support of our rounding algorithm, $A_h \geq \psi \cdot B_h$ where B_h is the Hamming error of h (when compared to $\bar{x}$). On the other hand, one can also use the properties of the rounding algorithm to get a lower bound on $\mathbb{E}[B_h]$ in terms of the recovery error (i.e., Hamming error) of the LP solution. These bounds together imply the required upper bound on the recovery error of the LP solution. $\square$

We defer the complete proof and an alternate dual-based proof to Appendix C.

Theorem 5.2 shows that on a $(2, 1, \psi)$-expansion stable instance, small objective gap $\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle$ implies small distance $\|x_V - \bar{x}_V\|_1$ in solution space. Although this holds for any $x \in L^*(G)$, we will be interested in x that are LP solutions to an observed, corrupted version of the stable instance.

We now show that if the observed instance has a nearby stable instance, then the LP solution for the observed instance has small Hamming error. For any two instances $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$ on the same graph G, the metric between them $d(\hat{\theta}, \bar{\theta}) := \sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle|$.

Corollary 5.3 (LP solution is good if there is a nearby stable instance)**.** *Let $\hat{x}^{MAP}$ and $\hat{x}$ be the MAP and local LP solutions to an observed instance $(G, \hat{c}, \hat{w})$. Also, let $\bar{x}$ be the MAP solution for a latent $(2, 1, \psi)$-expansion stable instance $(G, \bar{c}, \bar{w})$. If $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$,*

$$\frac{1}{2}\|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi} + \frac{1}{2}\|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

We defer the proof of this corollary to Appendix C.

6 Generative model for noisy stable instances

In the previous section, we showed that if an instance $(G, \hat{c}, \hat{w})$ is "close" to a $(2, 1, \psi)$-expansion stable instance $(G, \bar{c}, \bar{w})$ (i.e., the LP solution $\hat{x}$ to $(G, \hat{c}, \hat{w})$ has good objective in $(G, \bar{c}, \bar{w})$), then $\|\hat{x} - \bar{x}\|$ is small, where $\bar{x}$ is the MAP solution to the stable instance. In this section, we give a natural generative model for $(G, \hat{c}, \hat{w})$, based on randomly corrupting $(G, \bar{c}, \bar{w})$, in which $\hat{x}$ has good objective in $(G, \bar{c}, \bar{w})$ with high probability. Together with the curvature result from the previous section (Theorem 5.2), this implies that the LP relaxation, run on the noisy instances $(G, \hat{c}, \hat{w})$, approximately recovers $\bar{x}$ with high probability.

We now describe our generative model for the problem instances, which starts with an arbitrary stable instance and perturbs it with random additive perturbations to the edge costs and node costs (of potentially varying magnitudes). The random perturbations reflect possible uncertainty in the edge costs and node costs of the Markov random field. We will assume the random noise comes from any distribution that is sub-Gaussian[1]. However, there is a small technicality: the edge costs need to be positive (node costs can be negative). For this reason we will consider truncated sub-Gaussian random variables for the noise for the edge weights. We define sub-Gaussian and truncated sub-Gaussian random variables in Appendix D.

Generative Model: We start with an instance $(G, \bar{c}, \bar{w})$ that is $(2, 1, \psi)$-expansion stable, and perturb the edge costs and node costs independently. Given any instance $(G, \bar{c}, \bar{w})$, an instance $(G, \hat{c}, \hat{w})$ from the model is generated as follows:

1. For all node-label pairs (u, i), $\hat{c}(u, i) = \bar{c}(u, i) + \tilde{c}(u, i)$, where $\tilde{c}(u, i)$ is sub-Gaussian with mean 0 and parameter $\sigma_{u,i}$.
2. For all edges (u, v), $\hat{w}(u, v) = \bar{w}(u, v) + \tilde{w}(u, v)$, where $\tilde{w}(u, v)$ is an independent r.v. that is $(-w(u, v), \gamma_{u,v})$-truncated sub-Gaussian with mean 0.
3. $(G, \hat{c}, \hat{w})$ is the observed instance.

By the definition of our model, the edge weights $\hat{w}(u, v) \geq 0$ for all $(u, v) \in E$. The parameters of the model are the unperturbed instance $(G, \bar{c}, \bar{w})$, and the noise parameters $\{\gamma_{u,v}, \sigma_{u,i}\}_{u,v \in V, i \in [k]}$. On the one hand, the above model captures a natural average-case model for the problem. For a fixed ground-truth solution $x^* : V \to [k]$, consider the stable instance (H, c, w) where $w_{uv}^* = 2$ for all u, v in the same clus-

---

ter (i.e., $x^*(u) = x^*(v)$) and $w_{uv}^* = 1$ otherwise; and with $c^*(u, i) = 1$ if $x^*(u) = i$, and $c^*(u, i) = 1 + \psi$ otherwise. The above noisy stable model with stable instance (H, c, w) generates instances that are reminiscent of (stochastic) block models, with additional node costs. On the other hand, the above model is much more general, since we can start with *any* stable instance (G, c, w).

With high probability over the random corruptions of our stable instance, the local LP on the corrupted instance approximately recovers the MAP solution $\bar{x}$ of the stable instance. The key step in the proof of this theorem is showing that, with high probability, the observed instance is close to the latent stable instance in the metric we defined earlier.

Lemma 6.1 ($d(\hat{\theta}, \bar{\theta})$ is small w.h.p.). *There exists a universal constant $c < 1$ such that for any instance in the above model, with probability at least $1 - o(1)$,*

$$\sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle| \leq c\sqrt{nk} \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4} \sum_{uv} \gamma_{u,v}^2}$$

Proof (sketch). For any *fixed* $x \in L^*(G)$, we can show that $|\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle|$ is small w.h.p. using a standard large deviations bound for sums of sub-Gaussian random variables. The main technical challenge is in showing that the supremum over all feasible solutions is small w.h.p. The standard approach is to perform a union bound over an ε-net of feasible LP solutions in L^*. However, this gives a loose bound. Instead, we upper bound the supremum by using a rounding algorithm for LP solutions in $L^*(G)$, and union bound only over the discrete solutions output by the rounding algorithm. This gives significant improvements over the standard approach; for example, in a d-regular graph with equal variance parameter γ_{uv}, this saves a factor of $\sqrt{d}$ apart from logarithmic factors in n. □

We defer the details to Appendix D. The above proof technique that uses a rounding algorithm to provide a deviation bound for a continuous relaxation is similar to the analysis of SDP relaxations for average-case problems (see e.g., Makarychev et al., 2013; Guédon & Vershynin, 2016). The above lemma, when combined with Theorem 5.2 gives the following guarantee.

Theorem 6.2 (LP solution is nearly persistent). *Let $\hat{x}$ be the local LP solution to the observed instance $(G, \hat{c}, \hat{w})$ and $\bar{x}$ be the MAP solution to the latent $(2, 1, \psi)$-expansion stable instance $(G, \bar{c}, \bar{w})$. With high probability over the random noise,*

$$\frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 \leq \frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\sum_{u,i} \sigma_{u,i}^2 + k^2 \sum_{uv} \gamma_{u,v}^2}$$

Proof. We know that for any feasible solution $x \in L(G), \langle \bar{\theta}, x \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$. Therefore, $\langle \bar{\theta}, \hat{x} \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$. Remember that we defined $d(\hat{\theta}, \bar{\theta})$ as $\sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle|$. Since $\hat{x}$ and $\bar{x}$ are both points in $L^*(G)$,

$$\langle \bar{\theta}, \hat{x} \rangle \leq \langle \hat{\theta}, \hat{x} \rangle + d(\hat{\theta}, \bar{\theta}) \leq \langle \hat{\theta}, \bar{x} \rangle + d(\hat{\theta}, \bar{\theta})$$
$$\leq \langle \bar{\theta}, \bar{x} \rangle + 2d(\hat{\theta}, \bar{\theta})$$

The first and third inequalities follow from the definition of $d(\hat{\theta}, \bar{\theta})$. The second inequality follows from the fact that $\hat{x}$ is the minimizer of $\langle \hat{\theta}, x \rangle$ over all $x \in L(G)$. Therefore, $0 \leq \langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle \leq 2d(\hat{\theta}, \bar{\theta})$. Using this in Theorem 5.2, we get $\frac{1}{2}\|\hat{x} - \bar{x}\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi}$. Lemma 6.1 then gives an upper bound on $d(\hat{\theta}, \bar{\theta})$ that holds w.h.p. $\square$

For a d-regular graph in the uniform setting, we get the following useful corollary:

Corollary 6.3 (MAP solution recovery for regular graphs). *Suppose we have a d-regular graph G with $\gamma_{u,v}^2 = \gamma^2$ for all edges (u, v), and $\sigma_{u,i}^2 = \sigma^2$ for all vertex-label pairs (u, i). Also, suppose only a fraction ρ of the vertices and η of the edges are subject to the noise. With high probability over the random noise,*

$$\frac{\|\hat{x}_V - \bar{x_V}\|_1}{2n} \leq \frac{2ck\sqrt{\rho\sigma^2 + \frac{\eta dk}{8}\gamma^2}}{\psi}$$

Note that when $\hat{x}$ is an integer solution, the left-hand-side is the fraction of vertices misclassified by $\hat{x}$.

7   Finding nearby stable instances

In this section, we describe an algorithm for efficiently finding $(2, 1, \psi)$-expansion stable instances that are "close" to an observed instance $(G, \hat{c}, \hat{w})$. This algorithm allows us to check whether we can plausibly model real-world instances as "corrupted" versions of a $(2, 1, \psi)$-expansion stable instance.

In addition to an observed instance $(G, \hat{c}, \hat{w})$, the algorithm takes as input a "target" labeling x^t. For example, x^t could be a MAP solution of the observed instance. Surprisingly, once given a target solution, this algorithm is efficient.

We want to search over costs c and weights w. The broad goal to solve the following optimization problem:

$$\begin{array}{ll} \underset{c \geq 0, w \geq 0}{\text{minimize}} & f(c, w) \quad\quad (3) \\ \text{subject to} & (G, c, w) \text{ is } (2, 1, \psi)\text{-expansion stable} \\ & \text{with MAP solution } x^t, \end{array}$$

where $f(c, w)$ is any convex function of c and w. In particular, we will use $f_1(c, w) = \|(c, w) - (\hat{c}, \hat{w})\|_1$ and

$f_2(c, w) = \frac{1}{2}\|(c, w) - (\hat{c}, \hat{w})\|_2^2$ for minimizing the L1 and L2 distances between to the observed instance. The output of this optimization problem will give the closest objective vector $(\bar{c}, \bar{w})$ for which the instance $(G, \bar{c}, \bar{w})$ is $(2, 1, \psi)$-expansion stable. If the optimal objective value of this optimization is 0, the observed instance $(G, \hat{c}, \hat{w})$ is $(2, 1, \psi)$-expansion stable.

There is always a feasible (c, w) for (3) (see Appendix E for a proof), but it may change many weights and costs. Next we derive an efficiently-solvable reformulation of (3). In the next section, we find that the changes to $\hat{c}$ and $\hat{w}$ required to find a $(2, 1, \psi)$-expansion stable instance are relatively sparse in practice.

Theorem 7.1. *The optimization problem* (3) *can be expressed as a convex minimization problem over a polytope described by $poly(n, m, k)$ constraints. When $f(c, w) = \|(c, w) - (\hat{c}, \hat{w})\|_1$,* (3) *can be expressed as a linear program.*

The instance (G, c, w) is $(2, 1, \psi)$-expansion stable if x^t is better than every expansion y of x^t in every $(2, 1, \psi)$-perturbation of (c, w). Let $\mathcal{E}$ be the set of all expansions of the target solution x^t. Then for all θ' within a $(2, 1, \psi)$-perturbation of $\theta = (c, w)$, we should have that $\langle \theta', x^t \rangle \leq \min_{y \in \mathcal{E}} \langle \theta', x^t \rangle$. It is enough to check the *adversarial* $(2, 1, \psi)$-perturbation θ_{adv}. This perturbation makes the target solution x^t as bad as possible. If x^t is better than all the expansions $y \in \mathcal{E}$ in this perturbation, it is better than all $y \in \mathcal{E}$ in every $(2, 1, \psi)$-perturbation (see Appendix E for a proof). We set $\theta_{adv} = (c_{adv}, w_{adv})$ as:

$$c_{adv}(u, i) = \begin{cases} c(u, i) + \psi & i = x^t(u), \\ c(u, i) & \text{otherwise.} \end{cases}$$

$$w_{adv}(u, v) = \begin{cases} w(u, v) & x^t(u) \neq x^t(u), \\ \frac{1}{2}w(u, v) & \text{otherwise.} \end{cases}$$

The target solution x^t is fixed, so $\langle \theta_{adv}, x^t \rangle$ is a linear function of the optimization variables c and w. For $\alpha \in [k]$, let $\mathcal{E}^\alpha$ be the set of α-expansions of x^t. Because $\mathcal{E} = \cup_{\alpha \in [k]} \mathcal{E}^\alpha$, we have $\langle \theta', x^t \rangle \leq \min_{y \in \mathcal{E}} \langle \theta', x^t \rangle$ if and only if $\langle \theta', x^t \rangle \leq \min_{y \in \mathcal{E}^\alpha} \langle \theta', x^t \rangle$ for all $\alpha \in [k]$. We can simplify the original optimization problem to:

$$\begin{array}{ll} \underset{c \geq 0, w \geq 0}{\text{minimize}} & f(c, w) \\ \text{subject to} & \langle \theta_{adv}, x^t \rangle \leq \min_{y \in \mathcal{E}^\alpha} \langle \theta_{adv}, y \rangle \quad \forall\, \alpha \in [k], \end{array}$$

θ_{adv} is a linear function of c, w as defined above. We now use the structure of the sets $\mathcal{E}^\alpha$ to simplify the constraints. The optimal value of $\min_{y \in \mathcal{E}^\alpha} \langle \theta_{adv}, y \rangle$ is the objective value of the best (w.r.t. θ_{adv}) α-expansion of x^t. The best α-expansion of a fixed solution x^t can be found by solving a minimum cut problem in

Table 1: Results from the output of (5) on three stereo vision instances. More details in Appendix F.

Instance	Costs changed	Weights changed	(normalized) Recovery error bound	$\|\hat{x}_V - \hat{x}_V^{MAP}\|_1/2n$
`tsukuba`	4.9%	2.3%	0.0518	0.0027
`venus`	22.5%	1.3%	0.0214	0.0016
`cones`	1.2%	2.1%	0.0437	0.0022

an auxiliary graph $G_{aux}^{x^t}(\alpha)$ whose weights depend on linearly on the objective θ_{adv}, and therefore depend linearly on our optimization variables (c, w) (Boykov et al., 2001, Section 4). Therefore, the optimization problem $\min_{y \in \mathcal{E}^\alpha} \langle \theta_{adv}, y \rangle$ can be expressed as a minimum cut problem. Because this minimum cut problem can be written as a linear program, we can rewrite each constraint as

$$\langle \theta_{adv}, x^t \rangle \leq \min_{z: A(\alpha)z = b(\alpha), z \geq 0} \langle \theta_{adv}, z \rangle, \qquad (4)$$

where $\{A(\alpha)z = b(\alpha), \ z \geq 0\}$ is the feasible region of the standard metric LP corresponding to the minimum cut problem in $G_{aux}^{x^t}(\alpha)$. The number of vertices in $G_{aux}^{x^t}(\alpha)$ and the number of constraints in $A(\alpha)z = b(\alpha)$ is poly(m, n, k) for all α. We now derive an equivalent linear formulation of (4) using a careful application of strong duality. The dual to the LP on the RHS is:

$$\underset{\nu}{\text{maximize}} \quad \langle b(\alpha), \nu \rangle, \ \text{s.t.} \ A(\alpha)^T \nu \leq \theta_{adv}.$$

Because strong duality holds for this linear program, we have that (4) holds if and only if there exists ν with $A(\alpha)^T \nu \leq \theta_{adv}$ such that $\langle \theta_{adv}, x^t \rangle \leq \langle b(\alpha), \nu \rangle$.

This is a linear constraint in (c, w, ν). By using this dual witness trick for each label $\alpha \in [k]$, we obtain:

$$\underset{c \geq 0, w \geq 0, \{\nu_\alpha\}}{\text{minimize}} \quad f(c, w) \qquad (5)$$

$$\text{subject to} \quad \langle \theta_{adv}, x^t \rangle \leq \langle b(\alpha), \nu_\alpha \rangle \quad \forall \ \alpha$$

$$A(\alpha)^T \nu_\alpha \leq \theta_{adv} \quad \forall \ \alpha.$$

The constraints of (5) are linear in the optimization variables (c, w) and ν_α. The dimension of θ_{adv} is $nk + mk^2$, so there are $k(nk + mk^2 + 1)$ constraints and $nk + m + \sum_\alpha |b(\alpha)| = poly(m, n, k)$ variables. Because minimization of the L1 distance $f_1(c, w)$ can be encoded using a linear function and linear constraints, (5) is a linear program in this case. It is clear from the derivation of (5) that it is equivalent to (3). This proves Theorem 7.1. This formulation (5) can easily be input into "off-the-shelf" convex programming packages such as Gurobi (Gurobi Optimization, 2020).

8 Numerical results

Table 1 shows the results of running (5) on real-world instances of MAP inference to find nearby $(2, 1, \psi)$-

expansion stable instances. We study *stereo vision* models using images from the Middlebury stereo dataset (Scharstein & Szeliski, 2002) and Potts models from Tappen & Freeman (2003). Please see Appendix F for more details about the models and experiments.

We find, surprisingly, that only relatively sparse changes are required to make the observed instances $(2, 1, \psi)$-expansion stable with $\psi = 1$. On these instances, we evaluate the recovery guarantees by our bound from Theorem 5.2 and compare it to the actual value of the recovery error $\|\hat{x} - \hat{x}^{MAP}\|_1/2n$. In all of our experiments, we choose the *target* solution x^t for (5) to be equal to the MAP solution $\hat{x}^{MAP}$ of the observed instance. Therefore, we find a $(2, 1, \psi)$-expansion stable instance that has the same MAP solution as our observed instance. The recovery error bound given by Theorem 5.2 is then also a bound for the recovery error between $\hat{x}$ and $\hat{x}^{MAP}$, because $\hat{x}^{MAP} = x^t$. On these instances, the bounds from our curvature result (Theorem 5.2) are reasonably close to the actual recovery value. However, this bound uses the property that $\hat{x}$ has good objective in the stable instance and so it is still a "data-dependent" bound in the sense that it uses an empirically observed property of the LP solution $\hat{x}$. In Appendix F, we show how to refine Corollary 5.3 to give non-vacuous recovery bounds that do not depend on $\hat{x}$.

9 Conclusion

We studied the phenomenon of near persistence of the local LP relaxation on instances of MAP inference in ferromagnetic Potts model. We gave theoretical results, algorithms (for finding nearby stable instances) and empirical results to demonstrate that even after a $(2, 1, \psi)$-perturbation stable instance is corrupted with noise, the solution to the LP relaxation is nearly persistent i.e., it approximately recovers the MAP solution. Our theoretical results imply that the local LP approximately recovers the MAP solution when the observed instance is close to a stable instance. Our empirical results suggest that real-world instances are very close to stable instances. These results together suggest a new explanation for the near-persistency of the solution of the local LP relaxation for MAP inference in practice.

Acknowledgments

References

Abbe, E. Community detection and stochastic block models. *Foundations and Trends® in Communications and Information Theory*, 14(1-2):1–162, 2018. ISSN 1567-2190. doi: 10.1561/0100000067.

Angelidakis, H., Makarychev, K., and Makarychev, Y. Algorithms for stable and perturbation-resilient problems. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 438–451, 2017.

Archer, A., Fakcharoenphol, J., Harrelson, C., Krauthgamer, R., Talwar, K., and Tardos, É. Approximate classification via earthmover metrics. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1079–1087. Society for Industrial and Applied Mathematics, 2004.

Bilu, Y. and Linial, N. Are stable instances easy? In *Innovations in Computer Science*, pp. 332–341, 2010.

Birchfield, S. and Tomasi, C. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998.

Boykov, Y., Veksler, O., and Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

Chekuri, C., Khanna, S., Naor, J. S., and Zosin, L. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '01, pp. 109–118, USA, 2001. ISBN 0898714907.

Dahlhaus, E., Johnson, D. S., Papadimitriou, C. H., Seymour, P. D., and Yannakakis, M. The complexity of multiway cuts. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pp. 241–251, 1992.

Globerson, A. and Jaakkola, T. S. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In *Advances in neural information processing systems*, pp. 553–560, 2008.

Globerson, A., Roughgarden, T., Sontag, D., and Yildirim, C. How hard is inference for structured prediction? volume 37 of *Proceedings of Machine Learning Research*, pp. 2181–2190, Lille, France, 07–09 Jul 2015. PMLR. URL http://proceedings.mlr.press/v37/globerson15.html.

Guédon, O. and Vershynin, R. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016.

Gurobi Optimization, L. Gurobi optimizer reference manual, 2020. URL http://www.gurobi.com.

Kleinberg, J. and Tardos, E. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. *Journal of the ACM (JACM)*, 49(5):616–639, 2002.

Kolmogorov, V. and Wainwright, M. J. On the optimality of tree-reweighted max-product message-passing. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pp. 316–323, 2005.

Kovtun, I. Partial optimal labeling search for a np-hard subclass of (max,+) problems. In *Joint Pattern Recognition Symposium*, pp. 402–409. Springer, 2003.

Lang, H., Sontag, D., and Vijayaraghavan, A. Optimality of approximate inference algorithms on stable instances. In *International Conference on Artificial Intelligence and Statistics*, pp. 1157–1166, 2018.

Lang, H., Sontag, D., and Vijayaraghavan, A. Block stability for map inference. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 216–225, 2019.

Makarychev, K., Makarychev, Y., and Vijayaraghavan, A. Sorting noisy data with partial information. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pp. 515–528. ACM, 2013.

Makarychev, K., Makarychev, Y., and Vijayaraghavan, A. Bilu–linial stable instances of max cut and minimum multiway cut. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pp. 890–906. SIAM, 2014.

McSherry, F. Spectral partitioning of random graphs. In *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, FOCS '01, pp. 529–, Washington, DC, USA, 2001. IEEE Computer Society. ISBN 0-7695-1390-5. URL http://dl.acm.org/citation.cfm?id=874063.875554.

Meshi, O., London, B., Weller, A., and Sontag, D. Train and test tightness of lp relaxations in structured prediction. *The Journal of Machine Learning Research*, 20(1):480–513, 2019.

Roughgarden, T. (ed.). *Beyond the Worst-Case Analysis of Algorithms.* Cambridge University Press, 2021. doi: 10.1017/9781108637435.

Rowland, M., Pacchiano, A., and Weller, A. Conditions beyond treewidth for tightness of higher-order lp relaxations. In *Artificial Intelligence and Statistics*, pp. 10–18, 2017.

Scharstein, D. and Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47 (1):7–42, 2002.

Shekhovtsov, A. Exact and partial energy minimization in computer vision. 2013.

Shekhovtsov, A., Swoboda, P., and Savchynskyy, B. Maximum persistency via iterative relaxed inference in graphical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(7):1668–1682, 2017.

Sontag, D. *Approximate Inference in Graphical Models using LP Relaxations.* PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2010.

Sontag, D., Globerson, A., and Jaakkola, T. Introduction to dual decomposition for inference. In *Optimization for Machine Learning.* MIT Press, 2011.

Swoboda, P., Shekhovtsov, A., Kappes, J. H., Schnörr, C., and Savchynskyy, B. Partial optimality by pruning for map-inference with general graphical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(7), 2016.

Tappen and Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 900–906 vol.2, 2003.

Thapper, J. et al. The power of linear programming for valued csps. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 669–678. IEEE, 2012.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Wainwright, M. J. and Jordan, M. I. *Graphical models, exponential families, and variational inference.* Now Publishers Inc, 2008.

Wainwright, M. J., Jaakkola, T. S., and Willsky, A. S. Map estimation via agreement on trees: message-passing and linear programming. *IEEE transactions on information theory*, 51(11):3697–3717, 2005.

Weller, A., Rowland, M., and Sontag, D. Tightness of lp relaxations for almost balanced models. In *Artificial Intelligence and Statistics*, pp. 47–55, 2016.

Beyond Perturbation Stability: Supplementary Material

A   Preliminaries Details

Claim A.1. *For Uniform Metric Labeling, we can assume $c(u, i) \geq 0$ and $w(u, v) > 0$ without loss of generality.*

Proof. For problem instances where some node costs are strictly negative, let $c_{\min}$ be the minimum value among all the node costs. Consider a new problem instance where we keep the edge costs the same, but set $c'(u, i) = c(u, i) + |c_{\min}|$ for all $u \in V$ and $i \in [k]$. This new problem instance has all non-negative node costs, and the optimization problem is equivalent, because we added the same constant for all solutions. This reformulation also does not affect the $(2, 1)$-expansion stability or $(2, 1, \psi)$-expansion stability of the instance.

Likewise, for problem instances where some edge weights are 0, let E_0 be the set of all edges with 0 edge weight. Consider a new problem instance with $E' = E \setminus E_0$, with $w(u, v)$ unchanged for $(u, v) \in E \setminus E_0$, and identical node costs. The MAP optimization problem remains the same, and the new instance $((V, E'), c, w)$ is equivalent: it has the same MAP solution, and satisfies the stability definitions if and only if the original instance does as well. □

Claim 3.3. *For a given graph G, every solution $x \in L(G)$ that minimizes $\langle \theta, x \rangle$ for some valid objective vector $\theta = (c, w)$ also belongs to $L^*(G)$. Further, all integer solutions in $L(G)$ also belong to $L^*(G)$.*

Proof of Claim 3.3. Recall the local LP:

$$\min_x \sum_{u \in V} \sum_i c(u, i) x_u(i) + \sum_{(u,v) \in E} w(u, v) \sum_{i \neq j} x_{uv}(i, j) \tag{6}$$

$$\text{subject to:} \sum_i x_u(i) = 1 \qquad \forall \, u \in V \tag{7}$$

$$\sum_i x_{uv}(i, j) = x_v(j) \qquad \forall \, (u, v) \in E, \; j \in [k] \tag{8}$$

$$\sum_j x_{uv}(i, j) = x_u(i) \qquad \forall \, (u, v) \in E, \; i \in [k] \tag{9}$$

$$x_{uv}(i, j) \in [0, 1] \qquad \forall \, (u, v), \; (i, j) \tag{10}$$

$$x_u(i) \in [0, 1] \qquad \forall \, u, \; i. \tag{11}$$

The feasible region defined by the above constraints is $L(G)$. $L^*(G) \subseteq L(G)$ is the set of points that satisfy the additional constraint that $x_{uv}(i, i) = \min(x_u(i), x_v(i))$ for all $(u, v) \in E$ and $i \in [k]$. For any feasible node variable assignments $\{x_u\}$, $L^*(G)$ is not empty: a simple flow argument[2] implies that the constraints (8), (9), and (10) are always satisfiable even when we set $x_{uv}(i, i) = \min(x_u(i), x_v(i))$. For all integer feasible solutions in $L(G)$, notice that $x_{uv}(i, j) = 1$ if $x_u(i) = 1$ and $x_v(j) = 1$ or 0 otherwise. Therefore, all integer solutions satisfy this additional constraint. Consider a θ where all edge weights are strictly positive. If x minimizes $\langle \theta, x \rangle$, x must

---

[2]For an edge (u, v), consider the bipartite graph $\tilde{G} = ((U, V), E)$, where $|U| = |V| = k$. We let $x_u(i)$ represent the supply at node i in U, and let $x_v(i)$ represent the demand at node j in V. Because x_u and x_v are both feasible, the total supply equals the total demand. E contains all edges between U and V, so we can send flow from $i \in U$ to $j \in V$ for any (i, j) pair. Let $x_{uv}(i, j)$ represent this flow, and set $x_{uv}(i, i) = \min(x_u(i), x_v(i))$. For every i, this either satisfies the demand at node V_i or exhausts the supply at node U_i. In each case, we can remove that satisfied/exhausted node from the graph. After this choice of $x_u(i, i)$, the total remaining supply equals the total remaining demand $(\sum_i x_u(i) - \min(x_u(i), x_v(i)) = \sum_i x_v(i) - \min(x_u(i), x_v(i)))$, all supplies and demands are nonnegative, and the remaining graph $\tilde{G}'$ is a complete bipartite graph (over fewer nodes). This implies that the flow constraints (8), (9), (10) are still feasible.

Node	Costs		
u	.5	∞	∞
v	1	0	∞
w	1	∞	0

Figure 3: $(2,1)$-expansion stable instance that is not $(2,1)$-stable. In the original instance (shown left), the optimal solution labels each vertex with label 1, for an objective of 2.5. The adversarial $(2,1)$-perturbation for this instance replaces all the edge weights of $1+\varepsilon$ with $(1+\varepsilon)/2$. In this perturbed instance, the optimal solution labels $(u,v,w) \to (1,2,3)$. This has a node cost of 0.5 and an edge cost of $(3+3\varepsilon)/2$, for a total of $2 + 3\varepsilon/2 < 2.5$. Since the original solution is not optimal in the perturbed instance, this instance is not $(2,1)$-perturbation stable. However, note that the only expansions of the original solution (which had all label 1) that have non-infinite objective are $(u,v,w) \to (1,2,1)$ and $(u,v,w) \to (1,1,3)$. These each have objective $2.5 + \varepsilon$, which is strictly greater than the perturbed objective of the original solution. Therefore, this instance is $(2,1)$-expansion stable.

pay the minimum edge cost consistent with its node variables $x_u(i)$. So if we fix the $x_u(i)$ portion of x, we know that the edge variables x_{uv} of x are a solution to:

$$\min_{x \in L(G)} \sum_{(u,v) \in E} w(u,v) \sum_{i \neq j} x_{uv}(i,j).$$

Notice that since we have fixed the node variables $x_u(i)$, there is no interaction between the x_{uv} variables across different edges. So we can minimize this objective by minimizing each individual term $w(u,v) \sum_{i \neq j} x_{uv}(i,j)$. Since $w_{uv} > 0$ for all edges, we need to minimize $\sum_{i \neq j} x_{uv}(i,j)$. Notice that for every edge $(u,v) \in E$, we get that $\sum_i \sum_j x_{uv}(i,j) = 1$ by substituting $x_u(i)$ in constraint 7 with $\sum_j x_{uv}(i,j)$ from constraint 9. Therefore $\sum_{i \neq j} x_{uv}(i,j) = 1 - \sum_i x_{uv}(i,i)$. Thus, minimizing $\sum_{i \neq j} x_{uv}(i,j)$ is the same as maximizing $\sum_i x_{uv}(i,i)$. And the maximizing choice for $x_{uv}(i,i) = \min(x_u(i), x_v(i))$ due to constraints 8 and 9. $\square$

B Expansion Stability details

Claim B.1. *An instance (G, w, c) is $(2,1)$-expansion stable iff the MAP solution $\bar{x}$ is strictly better than all its expansions in the adversarial perturbation θ_{adv}. That is, for all $x \in \mathcal{E}_{\bar{x}}, \langle \theta_{adv}, x \rangle > \langle \theta_{adv}, \bar{x} \rangle$ where θ_{adv} has the same node costs c but has weights $w_{adv}(u,v) = \begin{cases} \frac{1}{2}w(u,v) & \bar{x}(u) = \bar{x}(v) \\ w(u,v) & \bar{x}(u) \neq \bar{x}(v). \end{cases}$*

Proof. Consider $\theta' = (c, w')$, any valid $(2,1)$-perturbation of $\theta = (c, w)$ i.e. for every edge $(u,v) \in E, \frac{w(u,v)}{2} \leq w'(u,v) \leq w(u,v)$. For any valid labeling x, let E_x represent the edges cut by x. Then, for any x which is an expansion of $\bar{x}$ i.e. $x \in \mathcal{E}_{\bar{x}}$,

$$\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle = \sum_{u \in V} c(u, x(u)) - c(u, \bar{x}(u)) + \sum_{(u,v) \in E_x} w'(u,v) - \sum_{(u,v) \in E_{\bar{x}}} w'(u,v)$$

$$= \sum_{u \in V} c(u, x(u)) - c(u, \bar{x}(u)) + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w'(u,v)$$

$$= \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - w_{adv}(u,v) + \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w_{adv} - w'(u,v)$$

$$= \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle + \sum_{(u,v) \in E_x \setminus E_{\bar{x}}} w'(u,v) - \frac{w(u,v)}{2} + \sum_{(u,v) \in E_{\bar{x}} \setminus E_x} w(u,v) - w'(u,v)$$

Since w' is a valid $(2,1)$-perturbation, $w'(u,v) \geq w(u,v)/2$ and $w'(u,v) \leq w(u,v)$. Therefore, for any valid $(2,1)$-perturbation θ', we have

$$\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \geq \langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle.$$

If the instance is $(2,1)$-expansion stable, then certainly $\langle \theta_{adv}, x \rangle > \langle \theta_{adv}, \bar{x} \rangle$ for all $x \in \mathcal{E}^{\bar{x}}$, since θ_{adv} is a valid $(2,1)$-perturbation of θ. If the instance is not $(2,1)$-expansion stable, there exists a θ' and an $x \in \mathcal{E}^{\bar{x}}$ for which

$\langle \theta', x \rangle - \langle \theta', \bar{x} \rangle \leq 0$. But the above inequality then implies that $\langle \theta_{adv}, x \rangle - \langle \theta_{adv}, \bar{x} \rangle \leq 0$ as well. This gives both directions. $\qquad \square$

This claim shows that to check whether an instance is $(2,1)$-expansion stable, it is sufficient to check that the MAP solution is strictly better than all its expansions in the adversarial perturbation θ_{adv}. We don't need to verify that this condition is satisfied in *every* valid $(2,1)$-perturbation. Because the optimal expansion of $\bar{x}$ in the instance with objective θ_{adv} can be computed efficiently, this claim also implies that $(2,1)$-expansion stability can be efficiently checked once the MAP solution $\bar{x}$ is known.

Claim B.2. $(2,1)$-*expansion stability is strictly weaker than* $(2,1)$-*perturbation stability.*

Proof. Figure 2 gives an instance of uniform metric labeling that is $(2,1)$-expansion stable but not $(2,1)$-perturbation stable. Here, $0 < \varepsilon < 1/3$. $\qquad \square$

Theorem 4.2 (Local LP is tight on $(2,1)$-expansion stable instances)**.** *Let $\bar{x}$ and $\hat{x}$ be the MAP and local LP solutions to a $(2,1)$-expansion stable instance (G, c, w), respectively. Then $\bar{x} = \hat{x}$ i.e. the local LP is tight on (G, c, w).*

Proof. First, we note that for any $x \in L^*(G)$, the objective value of the local LP can be written in a form that depends only on the node variables x_V. The objective term corresponding to the edges

$$
\begin{aligned}
\sum_{(u,v) \in E} w(u,v) \sum_{i \neq j} x_{uv}(i,j) &= \sum_{(u,v) \in E} w(u,v) \left(\sum_{i,j} x_{uv}(i,j) - \sum_i x_{uv}(i,i) \right) \\
&= \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i x_{uv}(i,i) \right) = \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i \min(x_u(i), x_v(i)) \right) \\
&= \sum_{(u,v) \in E} w(u,v) \left(1 - \sum_i \left(\frac{x_u(i) + x_v(i)}{2} - \frac{|x_u(i) - x_v(i)|}{2} \right) \right) \\
&= \sum_{(u,v) \in E} w(u,v) \left(\frac{1}{2} \sum_i |x_u(i) - x_v(i)| \right)
\end{aligned}
$$

Here we used the definition of $L^*(G)$ and the facts that $\sum_i x_u(i) = 1$ for all (u, i) and $\sum_j x_{uv}(i,j) = x_u(i)$ for all $(u, v) \in E, \ i \in [k]$.

Thus, for any $x \in L^*(G)$, the objective of the local LP can be written as

$$
\sum_{u \in V} \sum_i c(u,i) x_u(i) + \sum_{(u,v) \in E} w(u,v) d(u,v)
$$

where $d(u,v) := \frac{1}{2} \sum_i |x_u(i) - x_v(i)|$. This is the objective function of another LP relaxation for uniform metric labeling called the "metric LP", which is equivalent to the local LP (Archer et al., 2004). Note that both $\bar{x}$ and $\hat{x}$ are in $L^*(G)$ by Claim 3.3. Therefore, the objective function can be written in the above form for both of them.

In the next section, we introduce a rounding algorithm and prove some guarantees for the random solutions h output by it. We then use these guarantees to show an upper bound on the expected cost of these random solutions in a perturbed instance of the problem. Finally, we use this upper bound to prove that $\hat{x} = \bar{x}$.

B.1 ε-close rounding:

Given any feasible solution $x \in L(G)$ and a valid labeling $\bar{x}$, we construct a related feasible solution x' which is ε-close to $\bar{x}$ in the ℓ_∞-norm i.e. $\|x' - \bar{x}\|_\infty \leq \varepsilon$:

$$
x' = \varepsilon x + (1 - \varepsilon)\bar{x}, \tag{12}
$$

where $\varepsilon < 1/k$ and we have identified the labeling $\bar{x} : V \to \mathcal{L}$ with its corresponding vertex of the marginal polytope (a vector in $\{0,1\}^{nk+mk^2}$). We consider the following rounding algorithm applied to x', which is a modified version of the ε-close rounding algorithm used in Lang et al. (2018):

Algorithm 1 ε-close rounding

1: Choose $i \in \{1, \ldots, k\}$ uniformly at random.
2: Choose $r \in (0, 1/k)$ uniformly at random.
3: Initialize labeling $h : V \to [k]$.
4: **for** each $u \in V$ **do**
5:   **if** $x'_u(i) > r$ **then**
6:     Set $h(u) = i$.
7:   **else**
8:     Set $h(u) = \bar{x}(u)$
9:   **end if**
10: **end for**
11: **Return** h

Lemma B.3 (Rounding guarantees). *Given any x' constructed using* (12), *the labeling h output by Algorithm 1 satisfies the following guarantees:*

$$\mathbb{P}\{\, h(u) = i \,\} = x'_u(i) \qquad\qquad \forall\ u \in V, i \in [k]$$
$$\mathbb{P}\{\, h(u) \neq h(v) \,\} \leq 2d(u,v) \qquad\qquad \forall\ (u,v) \in E : \bar{x}(u) = \bar{x}(v)$$
$$\mathbb{P}\{\, h(u) = h(v) \,\} = (1 - d(u,v)) \qquad\qquad \forall\ (u,v) \in E : \bar{x}(u) \neq \bar{x}(v),$$

where $d(u,v) = \frac{1}{2}\sum_i |x'_u(i) - x'_v(i)|$ is the edge separation of the constructed feasible point x'.

Proof of Lemma B.3 (rounding guarantees). First, fix $u \in V$ and a label $i \neq \bar{x}(u)$. We output $h(u) = i$ precisely when i is chosen and $0 < r < x'_u(i)$, which occurs with probability $\frac{1}{k}\frac{x'_u(i)}{1/k} = x'_u(i)$ (we used here that $x'_u(i) \leq \varepsilon < 1/k$ for all $i \neq \bar{x}(u)$). Now we output $h(u) = \bar{x}(u)$ with probability $1 - \sum_{j \neq \bar{x}(u)} \mathbb{P}\{\, h(u) = j \,\} = 1 - \sum_{j \neq \bar{x}(u)} x'_u(j) = x'_u(\bar{x}(u))$, since $\sum_i x'_u(i) = 1$. This proves the first guarantee.

For the second, consider an edge (u,v) not cut by $\bar{x}$, so $\bar{x}(u) = \bar{x}(v)$. Then (u,v) is cut by h when some $i \neq \bar{x}(u)$ is chosen and r falls between $x'_u(i)$ and $x'_v(i)$. This occurs with probability

$$\frac{1}{k} \sum_{i \neq \bar{x}(u)} \frac{\max(x'_u(i), x'_v(i)) - \min(x'_u(i), x'_v(i))}{1/k} = \sum_{i \neq \bar{x}(u)} |x'_u(i) - x'_v(i)| \leq 2d(u,v).$$

Finally, consider an edge (u,v) cut by $\bar{x}$, so that $\bar{x}(u) \neq \bar{x}(v)$. Here $h(u) = h(v)$ if some i, r are chosen with $r < \min(x'_u(i), x'_v(i))$. We have $r < \min(x'_u(i), x'_v(i))$ with probability $\frac{\min(x'_u(i), x'_v(i))}{1/k}$. Note that this is still valid if $i = \bar{x}(u)$ or $i = \bar{x}(v)$, since only one of those equalities can hold. So we get

$$\frac{1}{k} \sum_i \frac{\min(x'_u(i), x'_v(i))}{1/k} = \frac{1}{2}\left(\sum_i x'_u(i) + x'_v(i) - |x'_u(i) - x'_v(i)| \right) = 1 - d(u,v),$$

where we used again that $\sum_i x'_u(i) = 1$. $\qquad\qquad\square$

Given these rounding guarantees, we can relate the expected cost difference between h and $\bar{x}$ in a perturbation of the original instance to the cost difference between x and $\bar{x}$ in the original instance. We are only interested in the case when $x \in L^*(G)$ and so the objective function $f(x) = \sum_{u \in V} \sum_i c(u,i) x_u(i) + \sum_{(u,v) \in E} w(u,v) d(u,v)$.

B.2   Using the rounding guarantees

Lemma B.4. *Given an integer solution $\bar{x}$, a feasible LP solution $x \in L^*(G)$, and a random output h of Algorithm 1 on an input $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$, define*

$$w'(u, v) = \begin{cases} \frac{1}{2}w(u,v) & \bar{x}(u) = \bar{x}(v) \\ w(u,v) & \bar{x}(u) \neq \bar{x}(v) \end{cases}$$

and let $f'(y) = \sum_{u \in V}\sum_i c(u,i)y_u(i) + \sum_{(u,v) \in E} w'(u,v)d(y,u,v)$ be the objective in the instance with the original costs, but using weights w'. Here $d(y,u,v) = \frac{1}{2}\sum_i |y_u(i) - y_v(i)|$. Let $A_h := f'(h) - f'(\bar{x})$ be the difference in this perturbed objective between h and $\bar{x}$. Then,

$$\mathbb{E}[A_h] = \mathbb{E}\left[f'(h) - f'(\bar{x})\right] \leq \varepsilon \cdot (f(x) - f(\bar{x})).$$

Proof.

$$\mathbb{E}\left[f'(h) - f'(\bar{x})\right] = \sum_{u \in V}\sum_i c(u,i)\mathbb{P}\left\{h(u) = i\right\} - \sum_{u \in V} c(u, \bar{x}(u)) + \sum_{uv:\bar{x}(u)=\bar{x}(v)} w'(u,v)\mathbb{P}\left\{h(u) \neq h(v)\right\}$$

$$- \sum_{uv:\bar{x}(u)\neq\bar{x}(v)} w'(u,v)\mathbb{P}\left\{h(u) = h(v)\right\}$$

$$= \sum_u\sum_i c(u,i)x'_u(i) - \sum_u c(u, \bar{x}(u)) + \sum_{uv:\bar{x}(u)=\bar{x}(v)} 2w'(u,v)d(x',u,v)$$

$$- \sum_{uv:\bar{x}(u)\neq\bar{x}(v)} w'(u,v)(1 - d(x',u,v))$$

$$= \sum_u\sum_i c(u,i)x'_u(i) - \sum_u c(u, \bar{x}(u)) + \sum_{uv \in E} w(u,v)d(x',u,v) - \sum_{uv:\bar{x}(u)\neq\bar{x}(v)} w(u,v)$$

$$= f(x') - f(\bar{x}).$$

where the second-to-last equality used the definition of w' (note that w' is identical to the worst-case perturbation w_{adv} for $\bar{x}$). Because f is convex (in particular, $d(x,u,v)$ is convex in x), we have $f(x') \leq \varepsilon f(x) + (1 - \varepsilon)f(\bar{x})$. Therefore,

$$\mathbb{E}\left[f'(h) - f'(\bar{x})\right] \leq \varepsilon f(x) + (1 - \varepsilon)f(\bar{x}) - f(\bar{x}) = \varepsilon(f(x) - f(\bar{x})),$$

which is what we wanted. □

B.3   Final proof of Theorem 4.2:

Suppose the local LP solution $\hat{x}$ is not the same as the MAP solution $\bar{x}$ i.e. $\hat{x} \neq \bar{x}$. Consider $x' = \varepsilon\hat{x} + (1 - \varepsilon)\bar{x}$ where $0 < \varepsilon < 1/k$ (see equation (12)). Let h be the random integer solution output by using Algorithm 1 on x'. By Lemma B.4, we have

$$\mathbb{E}\left[f'(h) - f'(\bar{x})\right] \leq \varepsilon \cdot (f(\hat{x}) - f(\bar{x}))$$

We note that any solution h that we get from rounding x' is either $\bar{x}$ or an expansion move of $\bar{x}$. This is because we pick only a single label i in step 1 of Algorithm 1 and label all vertices u either i or $\bar{x}(u)$. Therefore, for the i picked in step 1, h is an i-expansion of $\bar{x}$ if $h \neq \bar{x}$.

$$\mathbb{E}\left[f'(h) - f'(\bar{x})\right] = \mathbb{E}\left[f'(h) - f'(\bar{x})|h \neq \bar{x}\right]\mathbb{P}[h \neq \bar{x}] + \mathbb{E}\left[f'(h) - f'(\bar{x})|h = \bar{x}\right]\mathbb{P}[h = \bar{x}]$$

$$= \mathbb{E}\left[f'(h) - f'(\bar{x})|h \neq \bar{x}\right]\mathbb{P}[h \neq \bar{x}]$$

Since (G, c, w) is a $(2,1)$-expansion stable instance, we know that $f'(h) > f'(\bar{x})$ when $h \neq \bar{x}$ since all h in the support of the rounding (other than $\bar{x}$) are expansion moves of $\bar{x}$ and we get f' by a valid $(2,1)$-perturbation of (G, c, w). Therefore, $\mathbb{E}\left[f'(h) - f'(\bar{x})|h \neq \bar{x}\right] > 0$. We also have that $\mathbb{P}[h \neq \bar{x}] > 0$ since we assumed that $\hat{x} \neq \bar{x}$. Therefore, $\mathbb{E}\left[f'(h) - f'(\bar{x})\right] > 0$. But we know that $f(\hat{x}) - f(\bar{x}) \leq 0$ since $\hat{x}$ is the minimizer of $f(x)$ among all feasible $x \in L(G)$. So Lemma B.4 implies $\mathbb{E}\left[f'(h) - f'(\bar{x})\right] \leq 0$. Thus we have a contradiction and so the local LP solution $\hat{x}$ has to be the same as the MAP solution $\bar{x}$.

□

C Stability and Curvature around MAP solution: details

Theorem 5.2. *Let (G, c, w) be a $(2, 1, \psi)$-expansion stable instance with MAP solution $\bar{x}$. Let $\theta = (c, w)$. Then for any $x \in L^*(G)$, the recovery error (see Def. 3.1) satisfies*

$$\frac{1}{2}\|x - \bar{x}\|_1 := \frac{1}{2}\|x_V - \bar{x}_V\|_1 \leq \frac{1}{\psi}|\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle|. \tag{2}$$

Here, we provide two proofs for this theorem, one deals directly with the local LP relaxation and the other uses the dual of the relaxation. The dual proof is more general than the primal proof as it works for all $x \in L(G)$, not just for those in $L^*(G)$.

C.1 Primal-based proof

Proof. For any $x \in L^*(G)$, consider a feasible solution x' which is ε-close to $\bar{x}$ constructed using Equation 12 i.e. $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$. Let h be the random solution output by Algorithm 1 on x'.

Lemma C.1 (Bound for $\mathbb{E}[B_h]$). *For any h in the support of the rounding of $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$, let us define B_h to be the number of vertices which it labels differently from $\bar{x}$. In other words, it is the number of vertices which are misclassified by h i.e. $B_h := \sum_{u \in V} \mathbb{1}[h(u) \neq \bar{x}(u)]$. Then,*

$$\mathbb{E}[B_h] = \varepsilon \sum_{u \in V} \frac{1}{2}\|x_u - \bar{x}_u\|_1$$

Proof.

$$\mathbb{E}[B_h] = \sum_{u \in V} \mathbb{E}[\mathbb{1}[h(u) \neq \bar{x}(u)]] = \sum_{u \in V} \mathbb{P}\{h(u) \neq \bar{x}(u)\} = \sum_{u \in V} 1 - \mathbb{P}\{h(u) = \bar{x}(u)\}$$

$$= \sum_{u \in V} 1 - x'_u(\bar{x}(u)) = \sum_{u \in V} 1 - (\varepsilon x_u(\bar{x}(u)) + (1 - \varepsilon)) = \sum_{u \in V} \varepsilon(1 - x_u(\bar{x}(u)))$$

$$= \varepsilon \sum_{u \in V} \frac{1}{2}\left(1 - x_u(\bar{x}(u)) + \sum_{i \neq \bar{x}(u)} x_u(i)\right) = \varepsilon \sum_{u \in V} \frac{1}{2}\|x_u - \bar{x}_u\|_1$$

Here, we used the fact that for all $u \in V$, $\bar{x}_u(\bar{x}(u)) = 1$ and $\bar{x}_u(i) = 0 \; \forall \; i \neq \bar{x}(u)$ and since x is a feasible solution to the LP, it satisfies $\sum_{i \neq \bar{x}(u)} x_u(i) = 1 - x_u(\bar{x}(u))$ for all $u \in V$. $\qquad \square$

Lemma C.2 (Lower bound for A_h using $(2, 1, \psi)$-expansion stability). *If (G, w, c) is a $(2, 1, \psi)$-expansion stable instance, then for any h in the support of the rounding of $x' = \varepsilon x + (1 - \varepsilon)\bar{x}$,*

$$A_h \geq \psi \cdot B_h$$

where $A_h = f'(h) - f'(\bar{x})$ and f' is the objective in the instance (G, c, w') where w' is the worst $(2, 1)$ perturbation for $\bar{x}$ i.e.

$$w'(u, v) = \begin{cases} \frac{1}{2}w(u, v) & \bar{x}(u) = \bar{x}(v) \\ w(u, v) & \bar{x}(u) \neq \bar{x}(v) \end{cases}$$

Proof. Note that A_h here is the same as the one defined for Lemma B.4. Since the instance (G, c, w) is $(2, 1, \psi)$-expansion stable, we know that (G, c', w) should be $(2, 1)$-expansion stable for all c' such that $c \leq c' \leq c + \psi \cdot \mathbf{1}$. Consider the worst c' for $\bar{x}$ i.e. $c'(u, i) = \begin{cases} c(u, i) + \psi & i = \bar{x}(u) \\ c(u, i) & i \neq \bar{x}(u) \end{cases}$. Let f'' be the objective in the instance (G, c', w'). As discussed in section B.3, we know that any $h \neq \bar{x}$ in the support of the rounding is an expansion move of $\bar{x}$. Therefore, for any $h \neq \bar{x}$ in the support of the rounding of x',

$$f''(h) - f''(\bar{x}) > 0 \implies \sum_{u \in V} c'(u, h(u)) - c'(u, \bar{x}(u)) + \sum_{(u,v):h(u) \neq h(v)} w'(u, v) - \sum_{(u,v):\bar{x}(u) \neq \bar{x}(v)} w'(u, v) > 0$$

$$\implies \sum_{u \in V} c(u, h(u)) - (c(u, \bar{x}(u)) + \psi \cdot \mathbb{1}[h(u) \neq \bar{x}(u)]) + \sum_{(u,v):h(u) \neq h(v)} w'(u, v) - \sum_{(u,v):\bar{x}(u) \neq \bar{x}(v)} w'(u, v) > 0$$

$$\implies f'(h) - f'(g) > \psi \cdot \sum_{u \in V} \mathbb{1}[h(u) \neq \bar{x}(u)] \implies A_h > \psi \cdot B_h.$$

This is true for all $h \neq \bar{x}$ in the support of the rounding of x'. When $h = \bar{x}$, we have $A_h = B_h = 0$. Therefore for all h in the support of the rounding of x', we have that $A_h \geq \psi \cdot B_h$. □

C.2 Final proof of Theorem 5.2:

We use the Lemmas B.4(upper bound for $\mathbb{E}[A_h]$), C.2(lower bound for A_h), and C.1(bound for $\mathbb{E}[B_h]$) to prove Theorem 5.2. For all h in the support of rounding of x, $A_h \geq \psi \cdot B_h$. Also,

$$\mathbb{E}[A_h] \leq \varepsilon \left(f(x) - f(\bar{x}) \right), \ \mathbb{E}[B_h] = \varepsilon \sum_{u \in V} \frac{1}{2} \|x_u - \bar{x}_u\|_1$$

Suppose that $\|x - \bar{x}\|_1 > \tau \cdot (f(x) - f(\bar{x}))$. Then,

$$\frac{\mathbb{E}[A_h]}{\mathbb{E}[B_h]} \leq \frac{f(x) - f(\bar{x})}{\sum_{u \in V} \frac{1}{2} \|x_u - \bar{x}_u\|_1} < \frac{2}{\tau}$$

But since $A_h \geq \psi \cdot B_h$ for every h in the rounding of x, we get that $\frac{\mathbb{E}[A_h]}{\mathbb{E}[B_h]} \geq \psi$.

Setting $\tau = \frac{2}{\psi}$, we get a contradiction and thus we get,

$$\frac{1}{2} \|x - \bar{x}\|_1 \leq \frac{1}{\psi} \cdot (f(x) - f(\bar{x})) = \frac{1}{\psi} \cdot (\langle \theta, x \rangle - \langle \theta, \bar{x} \rangle)$$

□

Corollary 5.3 (LP solution is good if there is a nearby stable instance)**.** *Let $\hat{x}^{MAP}$ and $\hat{x}$ be the MAP and local LP solutions to an observed instance $(G, \hat{c}, \hat{w})$. Also, let $\bar{x}$ be the MAP solution for a latent $(2, 1, \psi)$-expansion stable instance $(G, \bar{c}, \bar{w})$. If $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$,*

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

Proof of Corollary 5.3. First, we note that for the nearby stable instance, the MAP and the local LP solutions are the same due to Theorem 4.2. Therefore, for any feasible solution $x \in L(G)$, $\langle \bar{\theta}, x \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$. In particular, this implies that $\langle \bar{\theta}, \hat{x} \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$ and $\langle \bar{\theta}, \hat{x}^{MAP} \rangle \geq \langle \bar{\theta}, \bar{x} \rangle$ since $\hat{x}, \hat{x}^{MAP}$ are also feasible solutions. Remember that we defined $d(\hat{\theta}, \bar{\theta}) := \sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle|$. Therefore,

$$\langle \bar{\theta}, \hat{x} \rangle \leq \langle \hat{\theta}, \hat{x} \rangle + d(\hat{\theta}, \bar{\theta}) \leq \langle \hat{\theta}, \bar{x} \rangle + d(\hat{\theta}, \bar{\theta}) \leq \langle \bar{\theta}, \bar{x} \rangle + 2d(\hat{\theta}, \bar{\theta}).$$

The first and third inequalities hold due to the definition of $d(\hat{\theta}, \bar{\theta})$. The second inequality follows from the fact that $\hat{x}$ is the minimizer for $\langle \hat{\theta}, x \rangle$ among $x \in L(G)$. Thus, $0 \leq \langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle \leq 2d(\hat{\theta}, \bar{\theta})$. From Theorem 5.2, we get $\frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 \leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi}$. Thus,

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \leq \frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1$$

$$\leq \frac{2d(\hat{\theta}, \bar{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

□

C.3 Dual-based proof

Here we provide an alternate proof of the curvature result using the dual of the local LP relaxation. First, we show that the curvature bound is related to the *dual margin* of the instance. Then we show that $(2, 1, \psi)$-expansion stability implies that the dual margin is at least ψ. Throughout this section, we assume the local LP solution $\hat{x}$ is unique and integral (as guaranteed, for example, by $(2, 1)$-expansion stability), so $\hat{x} = \bar{x}$.

Relaxing the local LP's marginalization constraints in both directions for each edge, we obtain the following Lagrangian for the local LP:

$$L(\delta, x) = \sum_u \sum_i \left(\theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i) \right) x_u(i) + \sum_{uv} \sum_{ij} \left(\theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j) \right) x_{uv}(i, j)$$

where each x_u is constrained to be in the $(k-1)$-dimensional simplex, and each x_{uv} the $k^2 - 1$-dimensional simplex (i.e., the normalization constraints remain). There are no constraints on the dual variables δ. Observe that for any δ and any primal-feasible x, $L(\delta, x) = \langle \theta, x \rangle$. This gives rise to the *reparametrization* view: for a fixed δ, define $\theta_u^\delta(i) = \theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i)$, and $\theta_{uv}^\delta(i, j) = \theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j)$. Then $L(\delta, x) = \langle \theta^\delta, x \rangle$. This will allow us to define equivalent primal problems with simpler structure than the original. $L(\delta, x)$ also gives the dual function:

$$D(\delta) = \min_x L(\delta, x) = \sum_u \min_i \left(\theta_u(i) + \sum_{v \in N(u)} \delta_{uv}(i) \right) + \sum_{uv} \min_{i,j} \left(\theta_{uv}(i, j) - \delta_{uv}(i) - \delta_{vu}(j) \right).$$

A dual point δ is a dual *solution* if $\delta \in \operatorname{argmax}_{\delta'} D(\delta')$. Theorem 4.2 implies that the local LP has a unique, integral solution when the instance is $(2, 1, \psi)$-expansion stable. Sontag et al. (2011, Theorem 1.3) show that this implies the existence of a dual solution δ^* that is *locally decodable* at all nodes u: for each u, $\operatorname{argmin}_i \theta_u^{\delta^*}(i)$ is unique, and moreover, the edge and node dual subproblems agree:

$$\left(\operatorname{argmin}_i \theta_u^{\delta^*}(i), \operatorname{argmin}_j \theta_v^{\delta^*}(j) \right) \in \operatorname{argmin}_{i,j} \theta_{uv}^{\delta^*}(i, j). \tag{13}$$

In this case, the primal solution defined by "decoding" δ^*, $x(u) = \operatorname{argmin}_i \theta_u^{\delta^*}(i)$, is the MAP solution (Sontag et al., 2011).

For locally decodable δ^*, we define the *node margin* $\psi_u(\delta^*)$ at a node u as:

$$\psi_u(\delta^*) = \min_{i \neq \operatorname{argmin}_j \theta^{\delta^*}(j)} \theta^{\delta^*}(i) - \min_j \theta^{\delta^*}(j).$$

This is the difference between the optimal reparametrized node cost at u and the next-smallest cost. Local decodability of δ is the property that $\psi_u(\delta) > 0$ for every u.

Together with (13), the following lemma implies that we need only consider locally decodable dual solutions where the optimal primal solution pays zero edge cost.

Lemma C.3 (Dual edge "removal"). *Given a locally decodable dual solution δ, we can transform it to a locally decodable dual solution δ' that satisfies $\min_{i,j} \theta_{uv}^{\delta'}(i, j) = 0$ and has the same (additive) margin at every node.*

Proof. Fix an edge (u, v), and consider any pair i^*, j^* in $\operatorname{argmin}_{i,j} \theta_{uv}^\delta(i, j)$. Put $\theta_{uv}^\delta(i^*, j^*) = \theta_{uv}(i^*, j^*) + \varepsilon$ for $\varepsilon \in \mathbb{R}$. Now define $\delta'_{uv}(i) = \delta_{uv}(i) - \varepsilon$ for all i (or, equivalently, $\delta'_{vu}(j) = \delta_{vu}(j) - \varepsilon$ for all j). Because we changed $\theta_u^\delta(i)$ by a constant for each i, local decodability is preserved and the additive *margin of local decodability* is not changed. We incurred a change of $+\varepsilon$ in the dual objective of δ from the edge term $\min_{i,j} \theta_{uv}^{\delta'}(i, j)$, and a $-\varepsilon$ in the objective from the decrease in the node term $\min_i \theta_u^{\delta'}(i)$, so δ' is still optimal. We can repeat this process for every edge (u, v). $\square$

Lemma C.3 implies that when (x^*, δ^*) is a pair of primal/dual optima and δ^* is locally decodable, we can assume that $L(x^*, \delta^*) = \sum_u \theta_u^{\delta^*}(x_u^*)$, where we overload notation to define x_u^* to be the label for which $x_u^*(i) = 1$. That is, the primal optimum pays no edge cost in the problem reparametrized by the dual opt δ^*. Finally, Lemma

C.3 implies that we can always assume that $\theta_{uv}^{\delta^*}(i,j) \geq 0$ for all (u,v), (i,j). Therefore, if there is any locally decodable dual solution, and the primal LP solution is integral and unique, we may assume there exists a locally decodable dual solution δ such that $\bar{x}(u) = \mathrm{argmin}_i \, \theta^\delta(i)$, $(\bar{x}(u), \bar{x}(v)) \in \mathrm{argmin}_{i,j} \, \theta_{uv}^\delta(i,j)$, $\theta_{uv}^\delta(\bar{x}(u), \bar{x}(v)) = 0$, and $\theta_{uv}^\delta(i,j) \geq 0$.

Lemma C.4 (Dual margin implies curvature around $\bar{x}$)**.** *For an instance with objective θ and MAP solution $\bar{x}$, assume there exists a locally decodable dual solution δ such that $\bar{x}(u) = \mathrm{argmin}_i \, \theta^\delta(i)$, $(\bar{x}(u), \bar{x}(v)) \in \mathrm{argmin}_{i,j} \, \theta_{uv}^\delta(i,j)$, $\theta_{uv}^\delta(\bar{x}(u), \bar{x}(v)) = 0$, and $\theta_{uv}^\delta(i,j) \geq 0$. Additionally, let $\psi(\delta) = \min_u \psi_u(\delta)$ be the smallest node margin. Note that $\psi(\delta) > 0$ because δ is locally decodable. Then for any $x \in L(G)$,*

$$\frac{1}{2}\|x_V - \bar{x}_V\|_1 \leq \frac{\langle \theta, x - \bar{x} \rangle}{\psi(\delta)}$$

Proof. Let $\Delta = \langle \theta, x - \bar{x} \rangle$. Since x and $\bar{x}$ are both primal-feasible, we have $L(x, \delta) = \langle \theta, x \rangle$ and $L(\bar{x}, \delta) = \langle \theta, \bar{x} \rangle$. Therefore,

$$L(x, \delta) = L(\bar{x}, \delta) + \Delta. \tag{14}$$

Because $\theta^\delta(\bar{x}(u), \bar{x}(v)) = 0$ for all (u,v), we have

$$L(\bar{x}, \delta) = \sum_u \theta_u^\delta(\bar{x}(u)).$$

Additionally, because $\theta_{uv}^\delta(i,j) \geq 0$,

$$L(x, \delta) = \sum_u \sum_i \theta_u^\delta(i) x_u(i) + \sum_{uv} \sum_{ij} \theta_{uv}^\delta(i,j) x_{uv}(i,j) \geq \sum_u \sum_i \theta_u^\delta(i) x_u(i).$$

Combining the above two inequalities with (14) gives:

$$\sum_u \sum_i \theta_u^\delta(i) x_u(i) \leq \sum_u \theta_u^\delta(\bar{x}(u)) + \Delta \tag{15}$$

Because δ is locally decodable to $\bar{x}$, and the smallest node margin is equal to $\psi(\delta)$, we have that for every u, $\theta_u^\delta(\bar{x}(u)) + \psi(\delta) \leq \theta_u^\delta(i)$ for all $i \neq \bar{x}(u)$. The margin condition implies:

$$\sum_u \theta_u^\delta(\bar{x}(u)) x_u(\bar{x}(u)) + \sum_u \sum_{i \neq \bar{x}(u)} (\theta_u^\delta(\bar{x}(u)) + \psi(\delta)) x_u(i) < \sum_u \sum_i \theta_u^\delta(i) x_u(i),$$

and simplifying using $\sum_i x_u(i) = 1$ gives:

$$\sum_u \theta_u^\delta(\bar{x}(u)) + \psi(\delta) \sum_u \sum_{i \neq \bar{x}(u)} x_u(i) < \sum_u \sum_i \theta_u^\delta(i) x_u(i).$$

Plugging in to (15) gives:

$$\sum_u \sum_{i \neq \bar{x}(u)} x_u(i) < \frac{\Delta}{\psi(\delta)}.$$

The left-hand-side is precisely $\|x_V - \bar{x}_V\|_1/2$. □

Now we show that $(2, 1, \psi)$-expansion stability implies that there exists a locally decodable dual solution δ with dual margin $\psi(\delta) \geq \psi$.

Lemma C.5 (($2, 1, \psi$)-expansion stability gives a lower bound on dual margin)**.** *Let (G, c, w) be a $(2, 1, \psi)$-expansion stable instance with $\psi > 0$. Then there exists a locally decodable dual solution δ with dual margin $\psi(\delta) \geq \psi$.*

Proof. Define new costs c_ψ as

$$c_\psi(u, i) = \begin{cases} c(u, i) + \psi & \bar{x}(u) = i \\ c(u, i) & \text{otherwise.} \end{cases}$$

By definition, the instance (G, c_ψ, w) is $(2, 1)$-expansion stable (see Definition 4.1). Theorem 4.2 implies the pairwise LP solution is unique and integral on (G, c_ψ, w). This implies there exists a dual solution δ^0 that is locally decodable to $\bar{x}$. The only guarantee on the dual margin of δ^0 is that $\psi(\delta^0) > 0$. But note that δ^0 is also an optimal dual solution for (G, c, w), since its objective in that instance is the same as the objective of $\bar{x}$. But in that instance, the dual margin at every node is at least ψ, because $c_\psi(u, \bar{x}(u)) - c(u, \bar{x}(u)) = \psi$. So $\psi(\delta) \geq \psi$. $\qquad\square$

These two lemmas directly imply Theorem 5.2. This dual proof is slightly more general than the primal proof, since the curvature result applies to any $x \in L(G)$.

D   Details for Generative model

Definition D.1 (sub-Gaussians and (b, σ)-truncated sub-Gaussians). Suppose $b \in \mathbb{R}, \sigma \in \mathbb{R}_+$. A random variable X with mean μ is sub-Gaussian with parameter σ if and only if $\mathbb{E}[e^{\lambda(X-\mu)}] \leq \exp(\lambda^2\sigma^2/2)$ for all $\lambda \in \mathbb{R}$. The random variable X is (b, σ)-truncated sub-Gaussian if and only if X is supported in (b, ∞) and X is sub-Gaussian with parameter σ.

We remark that the above definition captures many well-studied families of bounded random variables e.g., Rademacher distributions, uniform distributions on an interval etc. We remark that a bounded random variable supported on $[-M, M]$ is also sub-Gaussian with parameter M. However in our setting, it needs to be truncated only on negative side, and the bound M will be much larger than the variance parameter σ; the bound is solely to ensure non-negativity of edge costs. A canonical example to keep in mind is a truncated Gaussian distribution. We use the following standard large deviations bound for sums of sub-Gaussian random variables (for details, refer to Thm 2.6.2 from Vershynin (2018)). Given independent r.v.s $X_1, X_2, \ldots, X_n$, with X_i drawn from a sub-Gaussian with parameter σ_i we have for $\mu = \sum_{i=1}^n \mathbb{E}[X_i]$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$,

$$\mathbb{P}\Big[\Big|\sum_{i=1}^n X_i - \mu\Big| \geq t\Big] \leq 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big). \tag{16}$$

Lemma 6.1 $(d(\hat{\theta}, \bar{\theta})$ is small w.h.p.). *There exists a universal constant $c < 1$ such that for any instance in the above model, with probability at least $1 - o(1)$,*

$$\sup_{x \in L^*(G)} |\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle| \leq c\sqrt{nk}\sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4}\sum_{uv} \gamma_{u,v}^2}$$

Proof. As discussed in section B, for any $x \in L^*(G)$, the objective of the local LP can be written as

$$\sum_{u \in V}\sum_i c(u, i)x_u(i) + \sum_{(u,v) \in E} w(u, v)d(u, v)$$

where $d(u, v) := \frac{1}{2}\sum_i |x_u(i) - x_v(i)|$. Let $\hat{f}(x) := \langle \hat{\theta}, x \rangle, \bar{f}(x) := \langle \bar{\theta}, x \rangle$. Then,

$$|\langle \hat{\theta}, x \rangle - \langle \bar{\theta}, x \rangle| = |\hat{f}(x) - \bar{f}(x)| = \Big|\sum_{u \in V}\sum_{i \in L} \tilde{c}(u, i)x_u(i) + \sum_{(u,v) \in E} \tilde{w}(u, v)d(u, v)\Big|$$

For any feasible LP solution x, consider the following rounding algorithm $\mathcal{R}$:

Algorithm 2 $\mathcal{R}$ rounding

1: **for** each $i \in \mathcal{L}$ **do**
2:   Choose $r_i \in (0, 1)$ uniformly at random.
3:   **for** each $u \in V$ **do**
4:     **if** $x_u(i) > r_i$ **then**
5:       $\mathcal{R}(x)_u(i) = 1$.
6:     **else**
7:       $\mathcal{R}(x)_u(i) = 0$
8:     **end if**
9:   **end for**
10: **end for**

Then, we have

$$\mathbb{E}[\hat{f}(\mathcal{R}(x)) - \bar{f}(\mathcal{R}(x))] = \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) \, \mathbb{E}[\mathbb{1}[\mathcal{R}(x)_u(i) = 1]] + \sum_{(u,v) \in E} \frac{\tilde{w}(u, v)}{2} \sum_i \mathbb{E}[\mathbb{1}[\mathcal{R}(x)_u(i) \neq \mathcal{R}(x)_v(i)]]$$

$$= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) \mathbb{P}[x_u(i) > r_i] + \sum_{(u,v) \in E} \frac{\tilde{w}(u, v)}{2} \sum_i \mathbb{P}[\min(x_u(i), x_v(i)) \leq r_i < \max(x_u(i), x_v(i))]$$

$$= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) x_u(i) + \sum_{(u,v) \in E} \frac{\tilde{w}(u, v)}{2} \sum_i |x_u(i) - x_v(i)|$$

$$= \sum_{u \in V} \sum_{i \in L} \tilde{c}(u, i) x_u(i) + \sum_{(u,v) \in E} \tilde{w}(u, v) d(u, v) = \hat{f}(x) - \bar{f}(x)$$

Therefore,

$$\sup_{x \in L^*(G)} |\hat{f}(x) - \bar{f}(x)| = \sup_{x \in L^*(G)} |\mathbb{E}[\hat{f}(\mathcal{R}(x)) - \bar{f}(\mathcal{R}(x))]| \leq \sup_{\hat{x}_V \in \{0,1\}^{nk}} |\hat{f}(\hat{x}_V) - \bar{f}(\hat{x}_V)|$$

Note that for all $x \in L^*(G)$, $\hat{f}(x)$ and $\bar{f}(x)$ only depend on the portion of x restricted to the vertices i.e. x_V. This is why we only need to look at $\hat{x}_v \in \{0, 1\}^{nk}$ for the last inequality.

For any fixed $\hat{x}_V \in \{0, 1\}^{nk}$, since $\tilde{w}(u, v), \tilde{c}(u, i)$ are all mean 0 and sub-Gaussian with parameters $\gamma_{u,v}, \sigma_{u,i}$, we have for any $t > 0$,

$$\mathbb{P}\left[|\hat{f}(\hat{x}_V) - \bar{f}(\hat{x})| > t\right] \leq 2 \exp\left(\frac{-t^2}{2\left(\sum_{u,i} \sigma_{u,i}^2 + k^2/4 \sum_{uv} \gamma_{u,v}^2\right)}\right)$$

Taking $t = c\sqrt{nk}\sqrt{\sum_{u,i} \sigma_{u,i}^2 + k^2/4 \sum_{uv} \gamma_{u,v}^2}$, we get that for any fixed $\hat{x}_V \in \{0, 1\}^{nk}$,

$$\mathbb{P}\left[|\hat{f}(\hat{x}) - \bar{f}(\hat{x})| > t\right] \leq 2 \exp\left(-c^2 nk\right)$$

Taking a union bound over $\{0, 1\}^{nk}$, we get that

$$\mathbb{P}\left[\sup_{\hat{x}_V \in \{0,1\}^{nk}} |\hat{f}(\hat{x}) - \bar{f}(\hat{x})| > t\right] \leq 2 \exp\left(nk \left(\log 2 - c^2\right)\right)$$

$\square$

Here, c needs to be greater than $\sqrt{\ln 2} \approx 0.83$ to get a high probability guarantee.

Corollary 6.3 (MAP solution recovery for regular graphs). *Suppose we have a d-regular graph G with $\gamma_{u,v}^2 = \gamma^2$ for all edges (u, v), and $\sigma_{u,i}^2 = \sigma^2$ for all vertex-label pairs (u, i). Also, suppose only a fraction ρ of the vertices and η of the edges are subject to the noise. With high probability over the random noise,*

$$\frac{\|\hat{x}_V - \bar{x_V}\|_1}{2n} \le \frac{2ck\sqrt{\rho\sigma^2 + \frac{\eta dk}{8}\gamma^2}}{\psi}$$

Proof. From Theorem 6.2, we have that, with high probability over the random noise

$$\frac{1}{2}\|\hat{x}_V - \bar{x}_V\|_1 \le \frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4}\sum_{uv} \gamma_{u,v}^2}$$

In this setting, this leads to

$$\frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\sum_{u,i} \sigma_{u,i}^2 + \frac{k^2}{4}\sum_{uv} \gamma_{u,v}^2} = \frac{2}{\psi} \cdot c\sqrt{nk} \cdot \sqrt{\rho nk\sigma^2 + \eta\frac{k^2}{4}\frac{nd}{2}\gamma^2} = \frac{2cnk\sqrt{\rho\sigma^2 + \frac{\eta dk}{8}\gamma^2}}{\psi}$$

since $|V| = n, |L| = k$, and $|E| = \frac{nd}{2}$.

$\square$

E   Algorithm for finding nearby stable instances details

Let $\bar{x}$ be a MAP solution, and let $\mathcal{E}^{\bar{x}}$ be the set of expansions of $\bar{x}$. We prove that an instance is $(2, 1, \psi)$-expansion stable if and only if $\langle\theta_{adv}, \bar{x}\rangle \le \langle\theta_{adv}, x\rangle$ for all $x \in \mathcal{E}^{\bar{x}}$. In other words, it is sufficient to check for stability in the *adversarial* perturbation for $\bar{x}$. This proves that we need not check every possible perturbation when finding a $(2, 1, \psi)$-expansion stable instance.

Claim E.1. *Let (G, c, w) be an instance of uniform metric labeling with MAP solution $\bar{x}$. Define:*

$$w_{adv}(u, v) = \begin{cases} \frac{1}{2}w_{uv} & \bar{x}(u) = \bar{x}(v) \\ w(u, v) & \bar{x}(u) \ne \bar{x}(v) \end{cases}$$

$$c_{adv}(u, i) = \begin{cases} c(u, i) + \psi & \bar{x}(u) = i \\ c(u, i) & \text{otherwise.} \end{cases}$$

Let θ_{adv} be the objective vector in the instance (G, c_{adv}, w_{adv}). Then

$$\langle\theta', \bar{x}\rangle \le \langle\theta', x\rangle$$

for all $(2, 1, \psi)$-perturbations θ' of θ and all $x \in \mathcal{E}^{\bar{x}}$ if and only if:

$$\langle\theta_{adv}, \bar{x}\rangle \le \langle\theta_{adv}, x\rangle$$

for all $x \in \mathcal{E}^{\bar{x}}$.

Proof. The proof is analogous to that of Claim B.1. If the instance is $(2, 1, \psi)$-expansion stable, then $\langle\theta', x\rangle - \langle\theta', \bar{x}\rangle \ge 0$ for all $(2, 1, \psi)$-perturbations θ' and all expansions x of $\bar{x}$. Because θ_{adv} is a valid $(2, 1, \psi)$-perturbation, this gives one direction. For the other, note that if the instance is not $(2, 1, \psi)$-expansion stable, there exists a θ' and an $x \in \mathcal{E}^{\bar{x}}$ for which $\langle\theta', x\rangle - \langle\theta', \bar{x}\rangle < 0$. A direct computation shows that $\langle\theta', x\rangle - \langle\theta', \bar{x}\rangle \ge \langle\theta_{adv}, x\rangle - \langle\theta_{adv}, \bar{x}\rangle$ for all $(2, 1, \psi)$-perturbations θ' of θ. Then we have $\langle\theta_{adv}, x\rangle - \langle\theta_{adv}, \bar{x}\rangle < 0$. $\square$

This claim justifies (5), which only enforces that $\bar{x}$ is at least as good as all of its expansions in θ_{adv}. The following claim implies that there is always a feasible point of (5) that makes modifications of bounded size to c and w.

Claim E.2. *Consider an instance (G, c, w) with a unique MAP solution $\bar{x}$. Let w' be defined as*

$$w'(u, v) = \begin{cases} w(u, v) & \bar{x}(u) \neq \bar{x}(v) \\ 2w(u, v) & \bar{x}(u) = \bar{x}(v), \end{cases}$$

and let c' be defined as

$$c'(u, i) = \begin{cases} c(u, i) - \psi & \bar{x}(u) = i \\ c(u, i) & \bar{x}(u) \neq i. \end{cases}$$

Then the instance (G, c', w') is $(2, 1, \psi)$-expansion stable with MAP solution $\bar{x}$.

Proof (sketch). The original MAP solution $\bar{x}$ is also the MAP solution to (G, c', w'). Then the original instance (G, c, w) is obtained from (G, c', w') by performing the adversarial $(2, 1, \psi)$-perturbation for $\bar{x}$ (see Claim E.1). Because $\bar{x}$ was the unique MAP solution to this instance, it has better objective than all of its expansions. Therefore, (G, c', w') is $(2, 1, \psi)$-expansion stable, by Claim E.1. □

(G, c', w') is a "nearby" stable instance to (G, c, w), but it requires changes to quite a few edges—every edge that is not cut by $\bar{x}$—and changes the node costs of every vertex. Surprisingly, the stable instances we found in Section 8 were much closer than (G, c', w')—that is, only sparse changes were required to transform the observed instance (G, c, w) to a $(2, 1, \psi)$-expansion stable instance.

F   Experiment details

In this section, we give more details for the numerical examples for which we evaluate our curvature bound from Theorem 5.2. We studied instances for stereo vision, where the input is two images taken from slightly offset locations, and the desired output is the disparity of each pixel location between the two images (this disparity is inversely proportional to the depth of that pixel). We used the models from Tappen & Freeman (2003) on three images from the Middlebury stereo dataset (Scharstein & Szeliski, 2002). In this model, G is a grid graph with one node corresponding to each pixel in one of the images (say, the one taken from the left), the costs $c(u, i)$ are set using the Birchfield-Tomasi matching costs (Birchfield & Tomasi, 1998), and the edge weights $w(u, v)$ are set as:

$$w(u, v) = \begin{cases} P \times s & |I(u) - I(v)| < T \\ s & \text{otherwise.} \end{cases}$$

Here $I(u)$ is the intensity of one of the images (again, say the left image) at pixel location u, and we set $(P, T, s) = (2, 50, 4)$. This is a Potts model. The `tsukuba`, `cones`, and `venus` images were `120 x 150`, `125 x 150`, and `383 x 434`, respectively. These models had $k = 7$, $k = 5$, and $k = 5$, respectively.

To generate Table 1, we ran the algorithm in (5) using Gurobi (Gurobi Optimization, 2020) for the L1 distance. For each observed instance $(G, \hat{c}, \hat{w})$, this output a nearby $(2, 1, \psi)$-stable instance $(G, \bar{c}, \bar{w})$. In all of our experiments, we used $\psi = 1$. Additionally, we always set the target MAP solution x^t in (5) to be equal to the observed MAP solution $\hat{x}^{MAP}$. To evaluate our recovery bound, we compared the objective of the observed LP solution $\hat{x}$ to the $\hat{x}^{MAP}$ in $(G, \bar{c}, \bar{w})$. That is, if $\bar{\theta}$ is the objective for $(G, \bar{c}, \bar{w})$, we computed $\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle = \langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle$, where the second equality is because we set the target solution x^t to be equal to $\hat{x}^{MAP}$, so $\hat{x}^{MAP} = \bar{x}$. Because $\psi = 1$, the difference between these two objectives is precisely the value of our curvature bound. In particular, Theorem 5.2 guarantees that

$$\frac{1}{2n} ||\hat{x}_V - \hat{x}_V^{MAP}||_1 \leq \frac{1}{n} \left(\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle \right).$$

The right-hand-side is shown for these instances in the "Recovery error bound" column of Table 1, and the true value of $\frac{1}{2n} ||\hat{x}_V - \hat{x}_V^{MAP}||_1$ (i.e., the true recovery error) is shown in the identically titled column of Table 1. On these instances, $\frac{1}{n} \left(\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \hat{x}^{MAP} \rangle \right)$ is close to 0, so our curvature bound "explains" a large portion of $\hat{x}$'s recovery of $\hat{x}^{MAP}$. These instances are close to $(2, 1, \psi)$-stable instances where $\hat{x}$ and $\hat{x}^{MAP}$ have close objective, and this implies by Theorem 5.2 that $\hat{x}$ approximately recovers $\hat{x}^{MAP}$.

However, this result relies on a property of the LP solution $\hat{x}$: that it has good objective in the stable instance discovered by the procedure (5). Compare this to Corollary 5.3, which only depends on properties of the observed

instance $\hat{\theta}$ and the stable instance $\bar{\theta}$ (in particular, some notion of "distance" between them). Given an observed instance $\hat{\theta}$ and stable instance $\bar{\theta}$, we can try to compute $d(\bar{\theta}, \hat{\theta})$ from Corollary 5.3 to give a bound that does not depend on $\hat{x}$. Unfortunately, this distance can be large, leading to a bound that can be vacuous (i.e., normalized Hamming recovery > 1). The following refinement of Corollary 5.3 gives much tighter bounds.

Corollary F.1 (LP solution is good if there is a nearby stable instance, refined). *Let $\hat{x}^{MAP}$ and $\hat{x}$ be the MAP and local LP solutions to an observed instance $(G, \hat{c}, \hat{w})$. Also, let $\bar{x}$ be the MAP solution for a latent $(2, 1, \psi)$-expansion stable instance $(G, \bar{c}, \bar{w})$. If $\hat{\theta} = (\hat{c}, \hat{w})$ and $\bar{\theta} = (\bar{c}, \bar{w})$, define*

$$d(\bar{\theta}, \hat{\theta}) := \sup_{x \in L^*(G)} \langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle - (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle).$$

Note that while we still use the name $d(\cdot, \cdot)$, evoking a metric, d is not symmetric. Then

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \le \frac{d(\bar{\theta}, \hat{\theta})}{\psi} + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

Proof. Note:

$$\frac{1}{2} \|\hat{x}_V - \hat{x}_V^{MAP}\|_1 \le \frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 + \frac{1}{2} \|\hat{x}_V^{MAP} - \bar{x}_V\|_1.$$

By the definition of d, for any $x \in L^*(G)$,

$$\langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle \le d(\bar{\theta}, \hat{\theta}) + (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle).$$

Now if we set $x = \hat{x}$, the LP solution to the observed instance, we have $\langle \hat{\theta}, \hat{x} \rangle - \langle \hat{\theta}, \bar{x} \rangle) \le 0$, so

$$\langle \bar{\theta}, \hat{x} \rangle - \langle \bar{\theta}, \bar{x} \rangle \le d(\bar{\theta}, \hat{\theta}).$$

Theorem 5.2 then implies $\frac{1}{2} \|\hat{x}_V - \bar{x}_V\|_1 \le d(\bar{\theta}, \hat{\theta})/\psi$, which gives the claim. $\square$

Given an observed instance $\hat{\theta}$ and a $(2, 1, \psi)$-expansion stable instance $\bar{\theta}$ output by (5) with $\bar{x}^{MAP} = \hat{x}^{MAP}$, we can upper bound $d(\bar{\theta}, \hat{\theta})$ by computing

$$d_{up}(\bar{\theta}, \hat{\theta}) := \sup_{x \in L(G)} \langle \bar{\theta}, x \rangle - \langle \bar{\theta}, \bar{x} \rangle - (\langle \hat{\theta}, x \rangle - \langle \hat{\theta}, \bar{x} \rangle),$$

which is a linear program in x because we relaxed $L^*(G)$ to $L(G)$. Corollary F.1 then implies that the recovery error of $\hat{x}$ is at most $d_{up}(\bar{\theta}, \hat{\theta})/\psi$, which we can compute. Table 2 shows the results of this procedure on two of the same instances from Table 1 in the "Unconditional bound" column. While the values of this bound are much larger than the "Curvature bound" of Theorem 5.2, they are much more theoretically appealing, since they only depend on the difference between $\hat{\theta}$ and $\bar{\theta}$ rather than on a property of the LP solution $\hat{x}$ to $\hat{\theta}$. For Table 2, we did a grid search for ψ over $\{1, \dots, 10\}$; $\psi = 4$ gave the optimal unconditional bound for both instances. The difference in ψ explains the slight differences between the other columns of Tables 1 and 2.

Table 2: Results from the output of (5) on two stereo vision instances. Curvature bound shows the bound obtained from Theorem 5.2, which depends on the observed LP solution $\hat{x}$. Unconditional bound shows the bound from the refined version of Corollary 5.3, which depends *only* on the observed instance and the stable instance. This "unconditional" bound explains a reasonably large fraction of the LP solution's recovery for these instances: because the instance is close to a stable instance, the LP solution approximate recovers the MAP solution.

Instance	Costs changed	Weights changed	Curvature bound	Unconditional bound	$\|\hat{x}_V - \hat{x}_V^{MAP}\|_1/2n$
tsukuba	4.9%	2.8%	0.0173	0.4878	0.0027
cones	2.81%	2.31%	0.0137	0.2819	0.0022