Supplementary Materials:
Global Connectivity Potentials for Random Field Models

Sebastian Nowozin
Max Planck Institute for Biological Cybernetics
Spemannstr. 38, 72076 Tübingen, Germany
sebastian.nowozin@tuebingen.mpg.de

Christoph H. Lampert Max Planck Institute for Biological Cybernetics
Spemannstr. 38, 72076 Tübingen, Germany
christoph.lampert@tuebingen.mpg.de

1. Introduction

The supplementary materials contain proofs and additional explanations omitted from the main paper.

2. Proofs

Proof to Lemma 1. Every single node k constitutes a connected subgraph. By setting $y_k = 1$, $y_h = 0$ for $h \neq k$ a feasible solution is obtained. All these solutions are affinely independent. Furthermore the empty graph is also a feasible subgraph. It follows that $\dim(Z) = |V|$, i.e. the connected subgraph polytope has full dimension. $\square$

Proof to Lemma 2. First, $y_i \geq 0$. For each i, we construct $|V|$ affinely independent points in C with $y_i = 0$. Fix i, then one solution is obviously $\boldsymbol{x} = \mathbf{0}$, the empty subgraph. Next, for all $p \neq i$, obtain one solution by setting only $y_p = 1$, and for all $j \neq p$ set $y_j = 0$. Clearly, $y_j = 0$ and the $|V| - 1$ solutions thus obtained are affinely independent. In total we have $|V|$ solutions with $y_i = 0$, thus $y_i \geq 0$ is facet-defining.

Second, $y_i \leq 1$. Again let i be arbitrary. We construct $|V|$ affinely independent points in C with $y_i = 1$. For this, set $y_i = 1$ and $y_j = 0$ for all $j \neq i$. This is obviously one solution. Now root a spanning tree in i and set one node k at a time to $y_k = 1$, respecting the order of the spanning tree, i.e. the subgraph selected all nodes j with $y_j = 1$ always remains a connected subgraph of the spanning tree. This constructs $|V| - 1$ solutions, all affinely independent. Adding the first solution yields $|V|$ solutions in total, completing the proof. $\square$

Proof to Theorem 2. First, the direction "is feasible" implying "is connected". Assume any given feasible $\boldsymbol{y}$ given, hence any $y_i \in \{0, 1\}$. If $\sum_i y_i \leq 1$, the resulting subgraph is trivially connected, hence assume $\sum_i y_i \geq 2$. For arbitrary $y_i = 1$, $y_j = 1$, $i \neq j$, assume i and j are not connected, that is $(i, j) \notin E$ and moreover there exist no path on G with all vertex variables being one. Trivially, we construct a vertex-separator set $S = \{k \in V : y_k = 0\}$ with $S \in \mathcal{S}(i, j)$. The removal of S from V must disconnect i and j, as $(i, j) \notin E$. However, by (3) we must have $y_i + y_j - \sum_{k \in S} y_k - 1 = 2 - 0 - 1 = 1 \leq 0$, which is clearly violated. Thus, feasibility implies connectedness. Second, the direction "is connected" implying "is feasible". Take any $y_i = 1$, $y_j = 1$, $i \neq j$, and i, j connected in G by a path starting at i and ending at j such that all intermediate nodes k satisfy $y_k = 1$. For all separators $S \in \mathcal{S}(i, j)$, at least one node t of this path must satisfy $t \in S$. Therefore $y_i + y_j - \sum_{k \in S} y_k - 1 \leq y_i + y_j - y_t - 1 = 0 \leq 0$ is satisfied. Thus any connected subgraph is feasible. $\square$

Proof to Theorem 3. We will prove this for any i,$j \in V$ by constructing $|V|$ affinely independent points in C which satisfy the inequality as equality. By [4, section 9.2.3] this shows that the inequality is facet-defining.

For $i, j \in V$ arbitrarily chosen, for any $S \in \bar{\mathcal{S}}(i, j)$, let $S = \{s_1, \dots, s_{|S|}\}$ be the set of nodes in the essential vertex-separator set. Further let S induce a partitioning of the graph into the set S, the connected subgraphs P_i, P_j, containing i and j, respectively, and the connected subgraphs P_s connected to exactly one $s \in S$ (if it is connected to more than one $s \in S$, remove all but one edge arbitrarily). This is shown in Figure 1.
First, we construct $|P_i| + |P_j|$ affinely independent solutions in C which satisfy the equality.

1. For the connected subgraph P_i, root a spanning tree in i. Set $y_i = 1$, $y_k = 0$, $\forall k \in P_i, k \neq i$. For each such

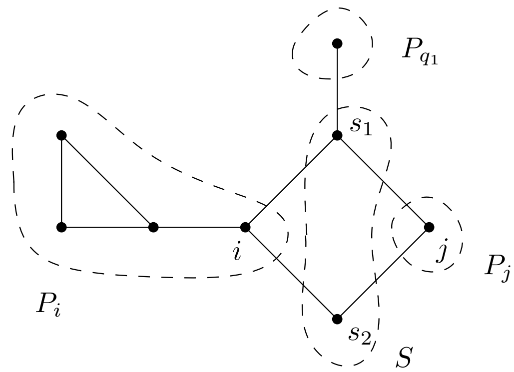

Figure 1. The separator set S induces a graph partitioning.

$k \in P_i$, enlarge the subgraph incrementally by one node in an arbitrary ordering respecting the spanning tree, i.e. set $y_k = 1$. Each enlarged solution is a connected subgraph of P_i and G, and affinely independent to all previous ones and satisfied the equality.

2. Likewise, do this for P_j, starting with just $y_j = 1$.

Next, for each $s \in S$, we construct $|P_s| + 1$ affinely independent solutions satisfying the equality as follows.

1. Set $y_k = 1, \forall k \in P_i \cup P_j$, and $y_s = 1$. This solution is in C because S is essential and thus s connects P_i and P_j. Construct $|P_s|$ more solutions by building a spanning tree for P_s, rooted in the node connected to s. By incrementally setting $y_k = 1$ in an order respecting the spanning tree, $|P_s|$ affinely independent solutions in C are obtained.

We now consider the total number of solutions constructed.

$$|P_i| + |P_j| + \sum_{s \in S} (|P_s| + 1) = |V|.$$

We have constructed $|V|$ affinely independent solutions in C satisfying the equality. Therefore, by [4, section 9.2.3], the inequality defines a facet of $\operatorname{conv}(C)$. □

3. Solution Integrality

In Section 4.1 we have evaluated the solution quality of the MRF with hard connectivity potential. Because we use relaxations for both the marginal polytope (the LP relaxation), and the connected subgraph polytope (the relaxation described by (5)), it is not a-priori clear that the solution obtained will be integral. Only if it is, we have a solution to the true, unrelaxed problem. If it is fractional, the solution is still optimal in the relaxation, but outside the true feasible set.

In Figure 2 we show the integrality, *i.e.* the fraction of variables which are integral. The reported numbers are the averages of 30 runs and the experimental setup is the same as in Section 4.1.

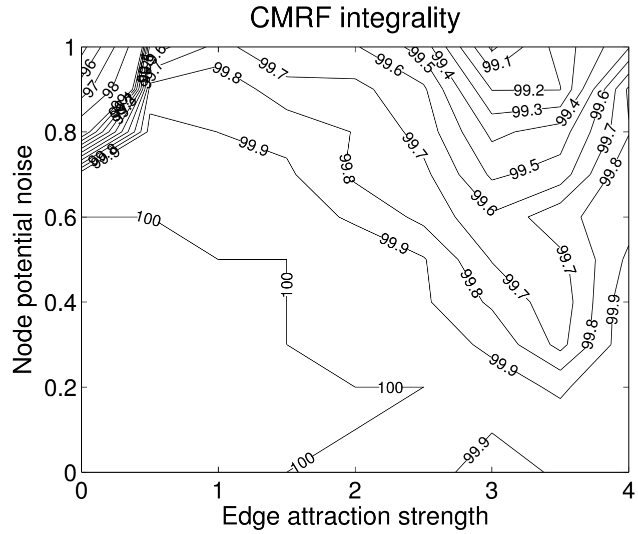

Figure 2. Mean solution integrality of the MRF with hard connectivity potential over 30 runs for varying problem parameters.

We see that our approach is very effective: for medium noise and edge interactions, the solution is always integral, whereas even when there is more noise and edge interaction, very few variables – less than 0.5% for most configurations – become fractional.

The problems defined by the marginal polytope and the connected subgraph polytope are both NP-hard. Hence, it is likely that no polynomial-time approach can provide the guaranteed optimum. In theory, a logical step within our approach would be to prove properties about the fractional solutions, for example that they satisfy half-integrality or can be rounded with optimality guarantees in order to obtain a polynomial-time approximation algorithm. In practise, the approach work already very well.

4. Implementation details

Separation routine. Our separation routine to find violated inequalities (5) is written in C++ and uses the boost 1.36 push-relabel maxflow solver.

MAP MRF linear program. We solve (2) using the open-source COIN-OR Clp 1.8 solver by using the COIN-OR Osi 0.98.2 interface.[1] Instead of generating a single constraint at a time, we use *multiple pricing* and add as many violated constraints as we can find in each iteration, usually a few thousand. The cost of re-solving the LP relaxation is small compared to generating constraints. Finding

[1]Clp is available at `https://projects.coin-or.org/Clp/`, Osi at `https://projects.coin-or.org/Osi/`, respectively.

additional violated constraints beside the most violating one incurs almost no additional cost.

Structured SVM. We solve (9) using the QP reformulation [3] in the dual by coordinate descent, similar to [2]. Unlike there, we need to ensure differentiability of the dual problem. Therefore, we add a small strictly convex proximal term in the primal, making it strictly convex in all variables. Strict convexity in the primal asserts dual differentiability everywhere [1], allowing our simple coordinate descent method to work. The advantage of the dual approach is the ability to rapidly warm-start once violating constraints have been found.

5. Experiments: Additional Results

Figure 3 confirms that if an object is present on an image, in 70% of the cases there is no other object of the same class on the image. For some classes, like `aeroplane`, `cat`, and `diningtable` this is more often the case than for classes like `bottle`, `chair`, `person` and `sheep`.

References

[1] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999. 2nd edition.

[2] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, and S. Sundararajan. A dual coordinate descent method for large-scale linear SVM. In *ICML*, volume 307, pages 408–415, 2008.

[3] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *JMLR*, 6:1453–1484, Sept. 2005.

[4] L. A. Wolsey. *Integer Programming*. John Wiley & Sons, New York, 1998.

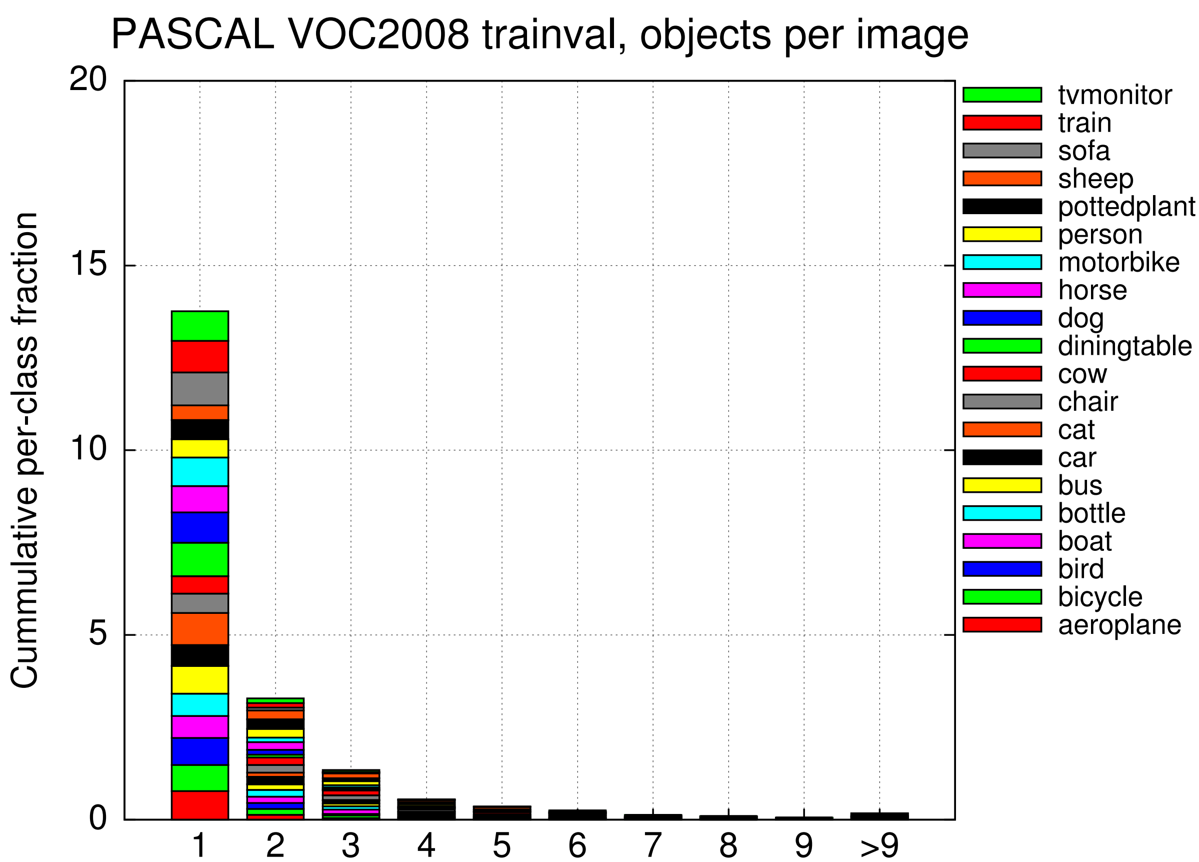

Figure 3. Number of objects of individual classes per image in the PASCAL VOC 2008 trainval data set.