MRF Inference by k-Fan Decomposition and Tight Lagrangian Relaxation

Jörg Hendrik Kappes, Stefan Schmidt, and Christoph Schnörr

Image and Pattern Analysis Group & HCI, University of Heidelberg, Germany,
{kappes,schmidt,schnoerr}@math.uni-heidelberg.de

Abstract. We present a novel dual decomposition approach to MAP inference with highly connected discrete graphical models. Decompositions into cyclic k-fan structured subproblems are shown to significantly tighten the Lagrangian relaxation relative to the standard local polytope relaxation, while enabling efficient integer programming for solving the subproblems. Additionally, we introduce modified update rules for maximizing the dual function that avoid oscillations and converge faster to an optimum of the relaxed problem, and never get stuck in non-optimal fixed points.

1 Introduction

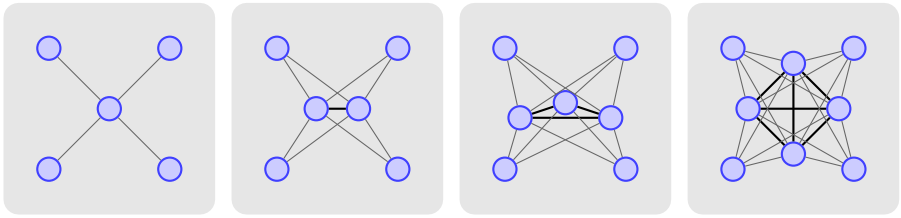
We focus on the Maximum A Posteriori (MAP) inference problem with discrete Markov Random Field (MRF) models. While applying graph cuts and iterated graph cuts has become standard for inference with exactly solvable submodular models, and for approximate inference with intractable models on sparse grid graphs, respectively [5, 21], recent research has focused on involved higher order models[1] [15, 9, 16], model decomposition and lower bound maximization based on linear programming (LP) duality [10, 12, 11, 17, 8], and tightening the common local polytope relaxation by advanced convex optimization [23, 24, 13, 20].

In this paper, we study the latter two points in connection with a particular class of highly connected graphical models, motivated by applications in computer vision. The models involve *k-fan substructures*[2] as subgraphs of the overall model. As illustrated in Figure 1, the defining property of this sub-structure is that an acyclic graph is obtained if we replace all inner nodes by a single node and merge resulting multiple edges.

Figure 2 illustrates our model for evaluating the HumanEva dataset [19]. Our graphical model detects the human pose in each image based on appearance features inside each view and epipolar-features between the 4 views. The random variables represent model parts (head, elbow, hand ...) defined over a finite set of image positions with a structure shown in Figure 2(b).

[1] The order of a model is given by the highest order of a term in the objective.

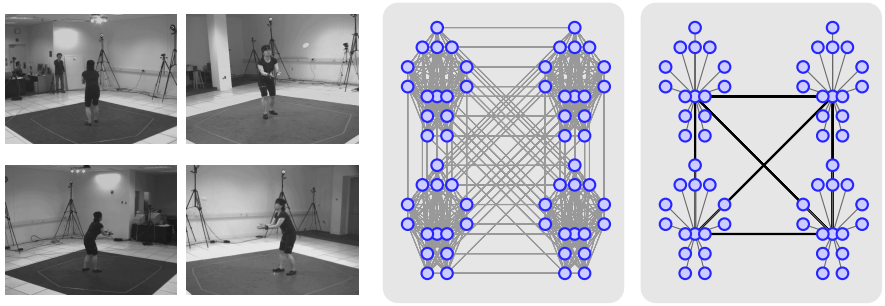
[2] We use the shorthand $G^{k,n}_{\text{fan}}$ for a fan graph with n nodes and k inner nodes. If n does not matter we just say k-fan.



(a) $G^{1,5}_{\text{fan}}$ (b) $G^{2,6}_{\text{fan}}$ (c) $G^{3,7}_{\text{fan}}$ (d) $G^{4,8}_{\text{fan}}$

Fig. 1. Examples for fan graphs. Inner nodes are connected to each other. Outer nodes are connected to all inner nodes, but not among each other.

Exact inference algorithm using this model is not feasible in acceptable time. By decomposing the problem in simple problems involving fan-structures as shown in Figure 2(c), however, high-quality inference becomes feasible by optimizing a bound on the relaxed linear problem via linear programming (LP) duality. We demonstrate below that utilizing fan-structures significantly improves the quality of the bounds obtained by standard LP relaxation.



(a) (b) (c)

Fig. 2. a) Images from the HumanEva dataset. Detection of human pose is done by processing all four views simultaneously, using the graphical model shown in **b)**. The model enhances a standard representation for each single view by epipolar constraints between views. **c)** shows one of 15 fan-structured subproblems used for tight relaxation.

The primary motivation for the decomposition of graphical models is twofold. Firstly, an approximation to the intractable inference problem can be formulated in terms of a two-level optimization procedure, where at the lower level inference on tractable substructures is carried out, while the master program at the upper level combines these partial solutions via dual variables. Secondly, the resulting objective value at the upper level yields a bound to the original (intractable) objective function, whose optimization through dual variables possibly meets the value of some locally computed optimum, thus providing a certificate that

this optimum is indeed a global one. For the general background, we refer to standard textbooks [3, 2], and for sophisticated applications of this principle in computer vision to, e.g. [10, 8, 12, 22].

A major difference to our work presented here is that we do *not* focus on graph decompositions into substructures for which inference can be efficiently done by standard methods, e.g. trees and belief propagation. While trees as substructures are natural for sparse grid graphs (e.g. row/column decomposition [10, 11]), they appear unnatural in connection with highly connected models as sketched above in Figure 2. Rather, we directly focus on more complex cyclic substructures, provided they are embedded into k-fans (Fig. 2(c)). This results in a relaxation of the overall inference problem that is provably tighter than the standard local polytope LP relaxation corresponding to tree structured subproblems. In the literature, cyclic substructures in connection with dual decomposition have been studied for the specific case of planar grid graphs by [11], in a general framework without specific evaluations by [8], and by [20] in terms of iteratively adding higher-order terms for locally improving the local polytope relaxation. Our work differs by focusing on k-fan substructures that can be flexibly applied to a wide range of non-planar, densely connected graphical models.

Another issue concerns the method for optimizing the Lagrangian dual at the upper level. In most work on dual decomposition of graphical models, convergence of the corresponding subgradient-based iteration is not really addressed. Either a "sufficiently small" step size is chosen, or the basic divergent series update rule is applied [7, 14]. In this paper, therefore, we merely raise this issue in the light of more recent pertinent work [18], due to its increasing importance in computer vision, leaving a more comprehensive investigation of this topic for future work.

Contribution. To summarize, our contribution consists in specifying mathematically the novel relaxation and showing both theoretically and empirically the influence of the choice of the subproblems on the relaxation and inference. In particular, we focus on k-fan substructures and general energy functions that are not restricted to any subclass. Furthermore, we improve convergence of the subgradient based optimization of the Lagrangian dual function.

Organization. We describe our problem decomposition and Lagrangian relaxation approach in Section 2 and show that this relaxation is tighter than the standard linear programming relaxation. Optimization of the Lagrangian dual via projected subgradient methods is discussed in Section 3. Finally, we present in Section 4 experimental results for synthetic and real world data.

Notation. Given a graph $G = (V, E)$ we associate to each node $a \in V$ a variable x_a taking values in $\mathcal{X}_a$ and a energy function $J(x) = \sum_{c \in C} f_c(x_c)$ with $C \subset V \cup E$. For $A \subset V$, we define $x_A = (x_a)_{a \in A}$ and $\mathcal{X}_A = \otimes_{a \in A} \mathcal{X}_a$, and as a shorthand $x = x_V$ and $\mathcal{X} = \mathcal{X}_V$. Following [23], we reformulate the problem of determining the optimal configuration x in $\mathcal{X}$,

$$x^* = \arg\min_{x \in \mathcal{X}} \sum_{c \in C} f_c(x_c), \tag{1}$$

in overcomplete form

$$\sum_{c \in C} f_c(x_c) = \langle \theta, \phi(x) \rangle = \sum_{i \in \mathcal{I}(G)} \theta_i \cdot \phi_i(x) \tag{2}$$

with vectors θ and $\phi(x)$ indexed by $\mathcal{I}(G) = \{(c; j) | c \in C,\ j \in \mathcal{X}_c\}$ and $\langle \cdot, \cdot \rangle$ denoting the inner product. Furthermore, given θ according to (2), problem (1) is equivalent to determining μ^* as solution to the LP

$$\mu^* = \arg\min_{\mu \in \mathcal{M}(G)} \langle \theta, \mu \rangle, \tag{3}$$

where $\mathcal{M}(G)$ denotes the marginal polytope defined as convex hull of all integer configurations with respect to the overcomplete representation. The exponentially large description of the feasible set $\mathcal{M}(G)$ reflects the combinatorial difficulty of the inference problem and necessitates problem approximations for general objective functions.

2 Problem Decomposition and Relaxation

In this section, it will be convenient to distinguish between original parameter vectors $\overline{\theta}, \overline{\theta}^i$ and parameter vectors θ^i defined by the problem decomposition – cf. (6) below. Starting with the convex optimization problem (3),

$$J(\mu^*) = \min_{\mu \in \mathcal{M}(G)} \langle \overline{\theta}, \mu \rangle, \tag{4}$$

we decompose it as follows. Given a set of graphs $\{G^1, \ldots, G^n\}$, with $G^i = (V, E^i)$ such that $E^i \subset E$ and $\bigcup_{i=1}^n E^i = E$, we define $\overline{\theta}^i \in \mathbb{R}^{\mathcal{I}(G)}$:

$$\overline{\theta}^i_{a;j} := \begin{cases} 0 & \text{if } a \notin V \cup E^i, \\ \overline{\theta}_{a;j}/n & \text{if } a \in V, \\ \overline{\theta}_{a;j}/\#a & \text{if } a \in E^i. \end{cases} \tag{5}$$

Here, $\#a$ denotes the number of edge-sets containing a. Note that the decomposition ensures $\overline{\theta} = \sum_i \overline{\theta}^i$. For each subproblem, we define another smaller exponential parameter vector

$$\theta^i := [\overline{\theta}^i]_{\mathcal{I}(G^i)} \tag{6}$$

called the *projection* of $\overline{\theta}^i$ with respect to $\mathcal{I}(G^i)$ and reformulate problem (4):

$$J(\mu^*) = \min_{\mu \in \mathcal{M}(G)} \sum_i \langle \overline{\theta}^i, \mu \rangle \tag{7a}$$

$$= \min_{\substack{\mu \in \mathcal{M}(G) \\ \forall i : \mu^i \in \mathcal{M}(G) \\ \forall i : \mu^i = \mu}} \sum_i \langle \overline{\theta}^i, \mu^i \rangle \overset{\text{eqn. (5)}}{=} \min_{\substack{\mu \in \mathcal{M}(G) \\ \forall i : \mu^i \in \mathcal{M}(G^i) \\ \forall i : \mu^i = [\mu]_{\mathcal{I}(G^i)}}} \sum_i \langle \theta^i, \mu^i \rangle \tag{7b}$$

$$\geq \min_{\substack{\mu \in \mathbb{R}^{\mathcal{I}(G)} \\ \forall i : \mu^i \in \mathcal{M}(G^i) \\ \forall i : \mu^i = [\mu]_{\mathcal{I}(G^i)}}} \sum_i \langle \theta^i, \mu^i \rangle \overset{\text{eqn. (5)}}{=} \min_{\substack{\mu \in \mathbb{R}^{\mathcal{I}(G)} \\ \forall i : [\mu]_{\mathcal{I}(G^i)} \in \mathcal{M}(G^i)}} \langle \overline{\theta}, \mu \rangle \tag{7c}$$

Decomposition (7) has the following properties:

- If all subgraphs are trees the relaxation is equivalent to the standard relaxation over the local polytope [23].
- If the subproblems include cycles, we get tighter relaxations which also take into account higher-order constraints.

In this paper, we focus on the latter option in terms of k-fan structured subproblems and show that this significantly tightens the relaxation and hence improves inference. Because problem (7c) is still difficult to solve, we focus on its dual by adding Lagrangian multipliers for the equality constraints, yielding the dual function

$$g(\lambda^1, \ldots, \lambda^n) := \min_{\substack{\mu \in \mathbb{R}^{\mathcal{I}(G)} \\ \forall i : \mu^i \in \mathcal{M}(G^i)}} \sum_i \langle \theta^i, \mu^i \rangle + \sum_i \sum_{\alpha \in \mathcal{I}(G^i)} \lambda^i_\alpha (\mu^i_\alpha - \mu_\alpha) \tag{8}$$

Since μ is unconstrained, this vector is determined by the corresponding partial derivatives of the right-hand side of (8). This yields the condition

$$(\lambda^1, \ldots, \lambda^n) \in \Lambda := \left\{ (\lambda^1, \ldots, \lambda^n) \;\middle|\; \forall \alpha \in \mathcal{I}(G) : \sum_{i \in \{j | \alpha \in \mathcal{I}(G^j)\}} \lambda^i_\alpha = 0 \right\}, \tag{9}$$

and by insertion into (8) the dual problem of the relaxed LP (7c)

$$\sup_{(\lambda^1, \ldots, \lambda^n) \in \Lambda} \sum_i \min_{\mu^i \in \mathcal{M}(G^i)} \langle (\theta^i + \lambda^i), \mu^i \rangle. \tag{10}$$

Since the feasible set of the primal problem (7c) includes a strict feasible point, Slater's condition [4] holds and guarantees that the duality gap between (7c) and its dual problem (10) is zero, i.e.

$$\sup_{(\lambda^1, \ldots, \lambda^n) \in \Lambda} g(\lambda^i, \ldots, \lambda^n) =: L^* \quad = \quad U^* := \min_{\substack{\mu \in \mathbb{R}^{\mathcal{I}(G)} \\ \forall i : [\mu]_{\mathcal{I}(G^i)} \in \mathcal{M}(G^i)}} \langle \overline{\theta}, \mu \rangle. \tag{11}$$

Instead of solving the relaxed primal problem (7c) which is still fairly complex, we can now solve the dual problem (10) by projected sub-gradient descent [3, 18], taking advantage of the problem decomposition into tractable subproblems. To this end, we have to optimize each subproblem

$$\min_{\mu^i \in \mathcal{M}(G^i)} \left\langle (\theta^i + \lambda^i), \mu^i \right\rangle \tag{12}$$

for a given λ^i (cf. section 3). Rather than solving the LP in (12) directly, we solve instead the corresponding integer programming problem. This is correct because vertices of the polytopes $\mathcal{M}(G^i)$ correspond to integer configurations. Accordingly, if the decomposition has been chosen properly, these integer problems can be solved very fast. As a by-product, we obtain an upper bound $\overline{U}(t)$ of the original objective function (4) by evaluating[3] in each step t for all subproblems i the solutions $(\mu^i)^t$:

$$\overline{U}(t) = \min_{t'=1,\dots,t} \min_{i=1,\dots,n} \langle \overline{\theta}, [(\mu^i)^{t'}]_{\mathcal{I}(G)} \rangle \tag{13}$$

The lower bound, on the other hand, reads

$$L(t) = \max_{t'=1,\dots,t} \sum_{i=1,\dots,n} \langle \theta^i + (\lambda^i)^{t'}, (\mu^i)^{t'} \rangle. \tag{14}$$

It crucially depends on the problem decomposition and thus reflects the quality of the relaxation. Figure 3 further explains and illustrates the relation between the different bounds and optima.

3 Solving the Dual Problem

The dual problem (10) is a nonsmooth concave maximization problem with linear constraints. The main difference between most inference algorithm based on dual decomposition [8, 11, 23, 24], besides the decomposition itself, concerns the choice and the computation of updates of λ in each step. A standard solver for such problems is the Projected Sub-Gradient Method (PSGM) [3] that requires to compute a subgradient of g at λ. The set of all subgradients at λ is called the subdifferential at λ and denoted by $\partial g(\lambda)$. We perform inference with respect to all subproblems and select a subgradient from the set

$$\partial g^i(\lambda^i) = \partial \left(\min_{\mu^i \in \mathcal{M}(G^i)} \langle \theta^i + \lambda^i, \mu^i \rangle \right) \tag{15}$$

$$= \left\{ \nabla \langle \theta^i + \lambda^i, \mu^* \rangle \,\middle|\, \mu^* \in \operatorname*{arg\,min}_{\mu^i \in \mathcal{M}(G^i)} g^i(\lambda^i) \right\} \tag{16}$$

$$= \left\{ \mu^* \,\middle|\, \mu^* \in \operatorname*{arg\,min}_{\mu^i \in \mathcal{M}(G^i)} g^i(\lambda^i) \right\}. \tag{17}$$

[3] The main trick in (13) is that $(\mu^i)^{t'}$ is integer and both graphs (G and G^i) have the same node set V, consequently the projection $[\cdot]_{\mathcal{I}(G)}$ is well defined.

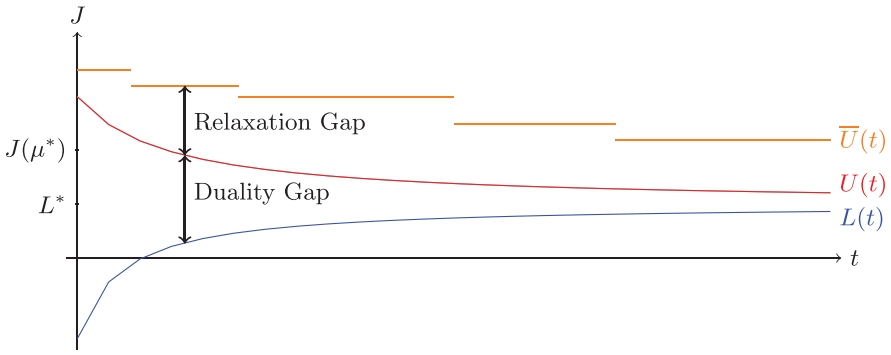


Fig. 3. This figure displays typical progressions of the bounds. Gray lines mark optimal values for the original primal and the relaxed primal/dual problem. Note that $L^* = U^*$ (zero duality gap for the relaxed problems). The blue line shows the current lower bound $L(t)$ of the dual relaxed problem. The red line marks the current upper bound $U(t)$ of the primal relaxed problem whose computation is too complex. Therefore we compute instead an upper bound $\overline{U}$ of the minimum of the original primal problem (marked in red). The observed gap $\overline{U}(t) - L(t)$ includes the current duality gap as well as the current relaxation gap, and we can not infer how they split up the total gap. However, we know that the duality gap will be zero and as a consequence, that after convergence the remaining gap is only due to the relaxation.

Concerning the subproblems, inference for a $G^{k,n}_{\text{fan}}$-structured model with L states per variable can be done using the junction tree algorithm [6] with asymptotic complexity $O((n-k) \cdot L^{(k+1)})$. We use an alternative search-based algorithm proposed in [1] having the same asymptotic worst case runtime complexity for fan graphs, but performs faster on average. For the synthetic data in section 4 this decreases the dominant term of the average complexity from L^{k+1} to approximately $L^{0.5 \cdot (k+1)}$, which is significant in practise.

Algorithm 1 shows a modified form of the PSGM. This modified version, also known as *heavy ball method*, does not step into the direction of the last subgradient, but rather into the direction of a convex combination of the subgradients observed so far. For $\rho^{(t)} = 1$, we obtain the standard PSGM, and for constant sequence $\rho^{(t)} \in (0, 1)$ a 'damped version' of it. We can guarantee that Algorithm 1 converges to an optimum provided that

$$\lim_{t \to \infty} \tau^{(t)} = 0, \qquad \sum_{t=0}^{\infty} \tau^{(t)} = \infty, \qquad \lim_{t \to \infty} \frac{\tau^{(t)}}{\rho^{(t)}} = 0. \tag{18}$$

For $\tau^{(t)} = \overline{\tau} \frac{1}{1+\alpha \cdot t}$ conditions (18) is satisfied for any constant sequence $\rho^{(t)}$. The speed of convergence depends highly on good choices of the sequence $\tau^{(t)}$, which we determine offline by grid-search on the parameter space for a particular problem class and the corresponding graphical model. However, a good choice

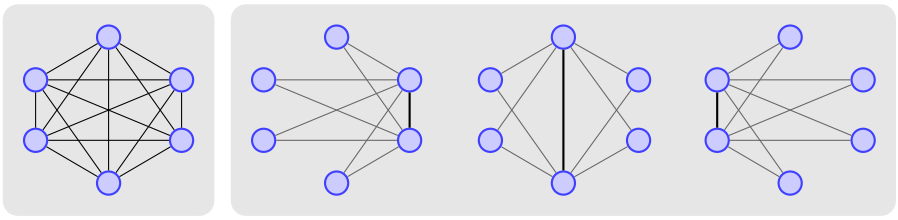
of $\rho^{(t)}$ also depends on the current value of $\tau^{(t)}$. Ruszczynski [18] suggests a damping sequence $\rho^{(t)} = \overline{\rho}\frac{\tau^{(t)}}{\overline{\tau}}$ and shows that for this sequence Algorithm 1 converges for $\overline{\rho} \in (0, 1]$, which generalizes the standard conditions (18).

Algorithm 1 Projected Sub-gradient Method

$t = 0, \quad \lambda^{(0)} = 0 \in \Lambda$
repeat
  $s \in \partial g((\lambda)^{(t)})$
  if $t == 0$ **then**
    $\zeta^{(t)} = s$
  else
    $\zeta^{(t)} = \zeta^{(t-1)} + \rho^{(t)}(s - \zeta^{(t-1)})$
  end if
  Compute $\overline{U}(t)$ and $L(t)$
  $(\lambda)^{(t+1)} = \left[(\lambda)^{(t)} + \tau^t \cdot \zeta^{(t)}\right]_\Lambda$
  $t = t + 1$
until $\|\overline{U}(t) - L(t)\| \leq \epsilon$ or $t > t_{\max}$

4 Experiments

Dual decomposition with k-fans: computer generated example. We first demonstrate the increasing accuracy achieved by selecting more complex k-fan subproblems, that is by raising k. We generated complete (fully connected) graphs $G = (V, E)$ containing first and second order potentials uniformly sampled between 0 and 1. Next we decomposed the graph into $\lceil |V|/k \rceil$ graphs $G^i = (V, E^i)$, as shown in Figure 4.



(a) G^6_{full} (b) Decomposition in three 2-fans

Fig. 4. A decomposition of a full connected graph with 6 nodes in three 2-fan structured graphs.

As Figure 5 reveals, we obtain much tighter bounds for decompositions with larger k. Also we achieved better integer solutions and this quite efficiently.

Decompositions into cyclic subproblems outperform decompositions with acyclic subproblems (lowest curve). This finding agrees with observations in [11, 8, 20]. Our results show that this property carries over to more complex problems with k-fan graphs as subproblems.

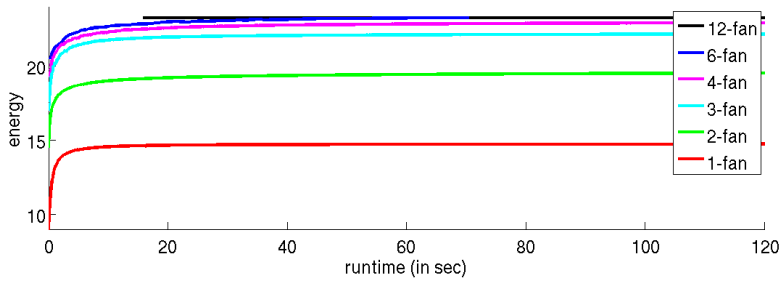


Fig. 5. Lower bounds for different k-fan decompositions as functions of runtime. The original problem involves 12 variables and 10 labels. The second order graph is fully connected, and the potentials were uniformly sampled between 0 and 1. Increasing k leads to tighter (larger) bounds.

Dual decomposition with k-fans: HumanEva dataset. The HumanEva dataset [19] contains 7 calibrated video sequences (4 grayscale and 3 color) that are synchronized together with 3D body poses obtained from a motion capture system. We used the 4 grayscale views and trained a model for images without taking into account temporal context. The graphical structure of our model is shown in Figure 2(b). Random variables take positions in the image domain, and the objective contains unary and pairwise potentials which include information about geometry and appearance. Nodes corresponding to each single view span a fully connected graph. Edges between views represent soft epipolar constraints. For a detailed description of the model, we refer to [1].

As inference for such models is difficult, we use a decomposition of this model into a set of 1-, 4- and 5-fans in order to derive a relaxation of the original problem as described above. The 1-fan corresponds to the approach of Komodakis [12] applied on our non-grid graphs. The 5-fan decomposition consist of 12 subproblems. Inner nodes of the fan of each subproblem correspond to the same single view, as sketched in Figure 7(c). Surprisingly the use of the 5-fans brings no advantage over the 1-fans for this problem. Rather, the decomposition with 1-fans gives even better results (see Table 1). The explanation for this is firstly, that the local relaxation inside a single view seems to be quite tight, and secondly that the computation of a subgradient of the 5-fan decomposition is more expensive than for 1-fans, hence takes more runtime. As we select *some* subgradient of the set $\partial g(\lambda)$, our current implementation does not check the optimality condition $0 \in g(\lambda)$ (Fermat condition), and therefore we additionally impose an upper bound on the runtime for possibly terminating the iteration.

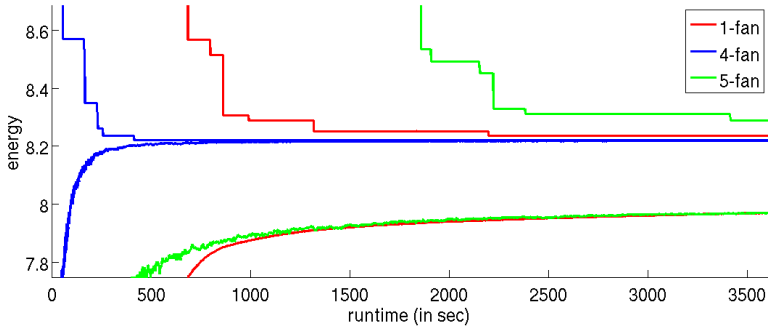
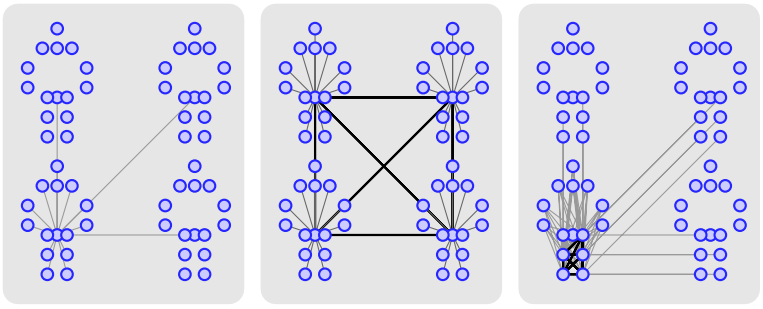


Fig. 6. The plot shows the progression of lower and upper bounds as a function of runtime. While both 1-fan and 5-fan decompositions restricted to single views perform similarly, the 4-fan decomposition enforcing epipolar consistency generates significantly tighter bounds and better integer solutions. The decompositions are sketched in Figure 7.

In order to effectively enforce higher order constraints between the same parts in different views, we set up a 4-fan decomposition in which the subproblems contain the clique of variables assigned to the same body-part in all views, as shown in Figure 7(b). This decomposition gives much better bounds and guarantees a gap which is less than $\epsilon = 10^{-6}$ in nearly 80% of the images and always outperforms the 1-fan decomposition. See Table 1 for more details.

These results show that while our approach can be applied to general graphical models, the overall performance may depend on the choice of a particular decomposition based on application-specific expertise.



(a) 1-fan (b) 4-fan (c) 5-fan

Fig. 7. The three graphs above sketch the structure of subproblems corresponding to three decompositions of the graphical model used for the HumanEva data. The 4-fan subgraphs include all epipolar constraints between single parts. The 5-fan decomposes the 15 nodes in each single view into three 5-fan substructures.

Table 1. We tested 3 different decompositions for 103 images from the HumanEva dataset. We decomposed the original problem into 1-fans, 4-fans and 5-fans as shown in Figure 7. We used a constant as well as a decreasing ρ-sequence (the latter one marked with *) for subgradient ascent with $\epsilon = 10^{-6}$. Choosing 5-fans inside single views does not improve inference while 4-fans between views results in much tighter relaxations. Furthermore, the decreasing ρ-sequence leads to faster convergence. The leftmost two columns specify how often the remaining gap was smaller than ϵ, and how often the best lower bound over all 6 approaches was reached. The 3rd and 4th column specify mean values of the gap and the lower bound. Finally, we compared the runtime for all data where all 6 approaches achieved ϵ-optimality.

	ϵ-gap convergence	best lower bound achieved	mean gap	mean lower bound	mean runtime
1-fan	27.18%	18.45%	0.0165	8.4816	1330 sec
1-fan*	35.92%	24.27%	0.0149	8.4820	1140 sec
4-fan	64.08%	60.19%	0.0011	8.4942	917 sec
4-fan*	78.64%	98.06%	0.0009	8.4943	577 sec
5-fan	11.65%	11.65%	0.0238	8.4760	2422 sec
5-fan*	23.30%	22.33%	0.0178	8.4805	1389 sec

5 Conclusions

We studied the decomposition of complex discrete graphical models into k-fan structured subproblems by Lagrangian relaxation. This enables to take into account more complex constraints as part of the subproblems that can still be solved to optimality within reasonable runtime. We also improved the performance of the subgradient ascent iteration for solving the Lagrangian dual problem, which is relevant not only for our problem but for any dual decomposition approach that are increasingly applied in computer vision research.

Experiments show that just choosing arbitrary decompositions into larger subproblems does not automatically lead to significantly better bounds. With little application-specific expertise, however, decompositions can be chosen that improve inference considerably, at moderate additional costs. The latter becomes immaterial for parallel implementations that are naturally supported by the problem decomposition.

Choosing automatically an optimal set of subproblems remains an open problem for future work, as is the case for automatically determining optimal parameter values for subgradient-based iterative optimization of nonsmooth dual functions.

Acknowledgement: Authors acknowledge partial support by the Research Training Group 1653 (http://graphmod.iwr.uni-heidelberg.de/) funded by the German Research Foundation (DFG).

References

[1] Bergtholdt, M., Kappes, J., Schmidt, S., Schnörr, C.: A study of parts-based object class detection using complete graphs. Int. J. Comp. Vision 87(1-2), 93–117 (2010)

[2] Bertsekas, D.: Nonlinear Programming. Athena Scientific, Belmont, Mass., 2nd edn. (1999)

[3] Bertsimas, D., Tsitsiklis, J.: Introduction to Linear Optimization. Athena Scientific (1997)

[4] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York, NY, USA (2004)

[5] Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Patt. Anal. Mach. Intell. 23(11), 1222–1239 (2001)

[6] Cowell, R.G., Dawid, A.P., Lauritzen, S.L., Spiegelhalter, D.J.: Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks. Springer Publishing Company, Incorporated (2007)

[7] Ermoliev, Y.: Methods for solving nonlinear extremal problems. Cybernetics 2(4), 1–17 (1966)

[8] Johnson, J.K., Malioutov, D., Willsky, A.S.: Lagrangian relaxation for MAP estimation in graphical models. In: 45th Annual Allerton Conference on Communication, Control and Computing (September 2007)

[9] Kohli, P., Ladický, L., Torr, P.H.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vision 82(3), 302–324 (2009)

[10] Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. 28(10), 1568–1583 (2006)

[11] Komodakis, N., Paragios, N.: Beyond pairwise energies: Efficient optimization for higher-order MRFs. In: CVPR. pp. 2985–2992. IEEE (2009)

[12] Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: ICCV. pp. 1–8. IEEE (2007)

[13] Kumar, M.P., Kolmogorov, V., Torr, P.H.S.: An analysis of convex relaxations for MAP estimation of discrete MRFs. J. Mach. Learn. Res. 10, 71–106 (2009)

[14] Polyak, B.: A general method for solving extremum problems. Soviet Math. 8, 593–597 (1966)

[15] Potetz, B., Lee, T.S.: Efficient belief propagation for higher-order cliques using linear constraint nodes. Comput. Vis. Image Underst. 112(1), 39–54 (2008)

[16] Rother, C., Kohli, P., Feng, W., Jia, J.: Minimizing sparse higher order energy functions of discrete variables. In: CVPR. pp. 1382–1389. IEEE (2009)

[17] Rother, C., Kolmogorov, V., Lempitsky, V.S., Szummer, M.: Optimizing binary MRFs via extended roof duality. In: CVPR. IEEE Computer Society (2007)

[18] Ruszczynski, A.: A merit function approach to the subgradient method with averaging. Optimization Methods Software 23(1), 161–172 (2008)

[19] Sigal, L., Black, M.: HumanEva: synchronized video and motion capture dataset for evaluation of articulated human motion. Tech. Rep. Technical Report CS-06-08, Brown University (2006)

[20] Sontag, D., Meltzer, T., Globerson, A., Jaakkola, T., Weiss, Y.: Tightening LP relaxations for MAP using message passing. In: McAllester, D.A., Myllymki, P. (eds.) UAI. pp. 503–510. AUAI Press (2008)

[21] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for markov random fields with smoothness-based priors. IEEE Trans. Patt. Mach. Intell. 30(6), 1068–1080 (2008)

[22] Vicente, S., Kolmogorov, V., Rother, C.: Joint optimization of segmentation and appearance models. In: Proc. ICCV'09 (2009)
[23] Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning 1(1-2), 1–305 (2008)
[24] Werner, T.: Revisiting the decomposition approach to inference in exponential families and graphical models. Tech. rep., Center for Machine Perception, Czech Technical University (May 2009)