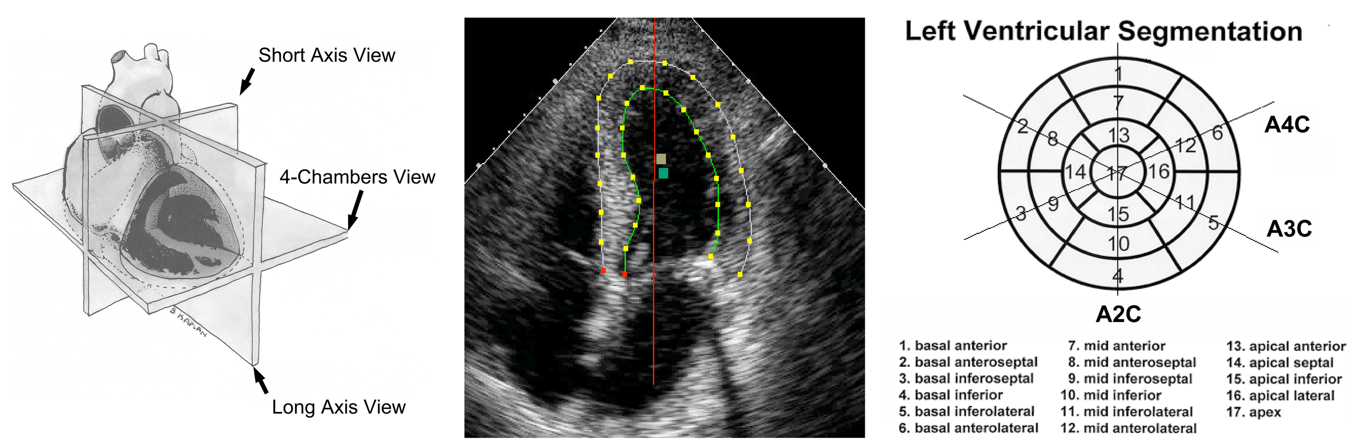
Structure Learning in Random Fields for Heart Motion Abnormality Detection

Mark Schmidt[1], Kevin Murphy[1], Glenn Fung[2], Romer Rosales[2]

[1]Dept. of Computer Science, University of British Columbia, [2]IKM CAD and Knowledge Solutions USA Inc., Siemens Medical Solutions

Introduction

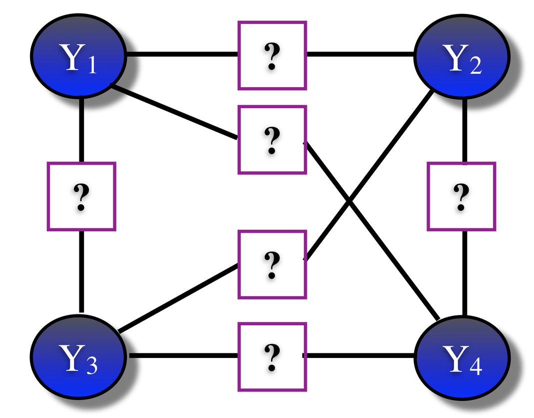
- We build a classifier that assists doctors in detecting *Coronary Heart Disease* from the motion of 16 left ventricle segments in ultrasound video



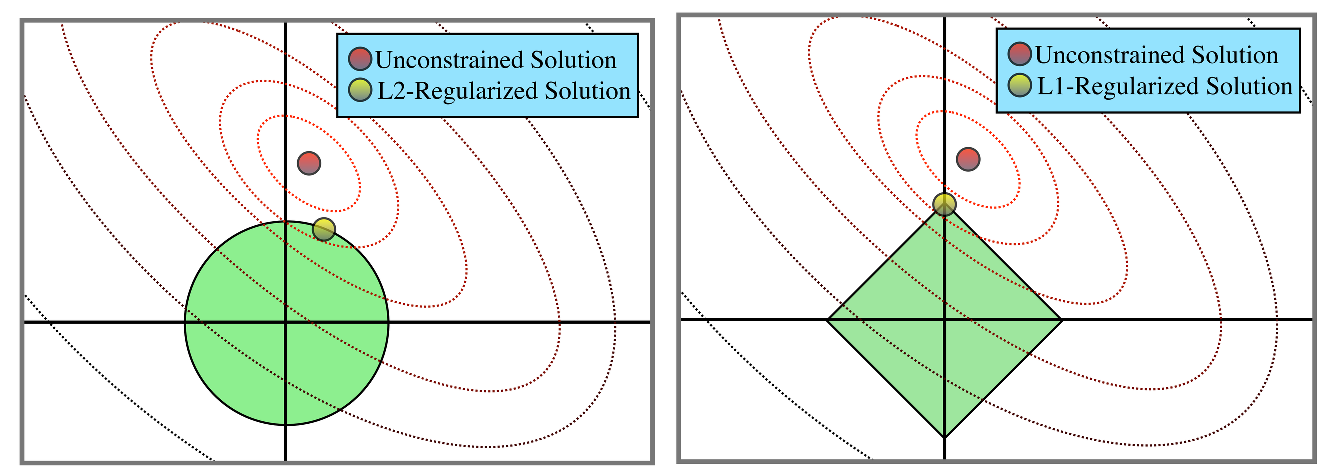
- *Conditional Random Fields* (CRFs) are used to model correlations between segments
- Our new *Group-L1 regularized optimization* algorithm lets us simultaneously learn the parameters and structure of CRFs

L1-Regularization for Structure Learning

- We want to learn the graph structure of the CRF labels

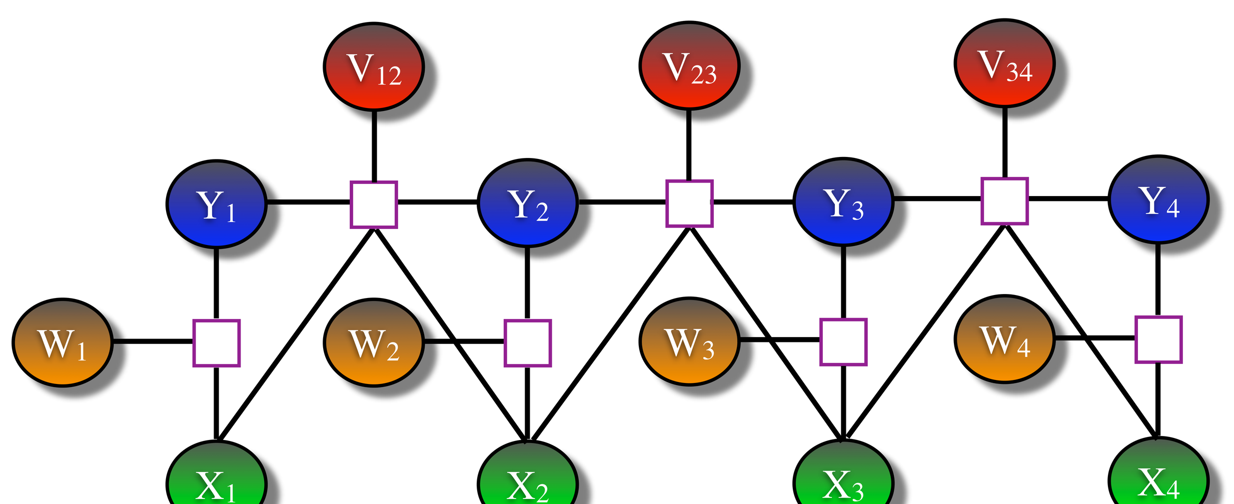


- L1 Regularization of edge parameters during learning leads to *sparsity*



- However, each edge has multiple parameters so we must consider *Group-L1 Regularization*

Conditional Random Fields



- Discriminative classifier modeling local and pairwise potentials of labels Y given data X

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\mathcal{Z}(\mathbf{x})} exp(\sum_i \mathbf{x}_i^T \mathbf{w}_{y_i}^i + \sum_{<ij>} \mathbf{x}_{i,j}^T \mathbf{v}_{y_i,y_j}^{i,j})$$

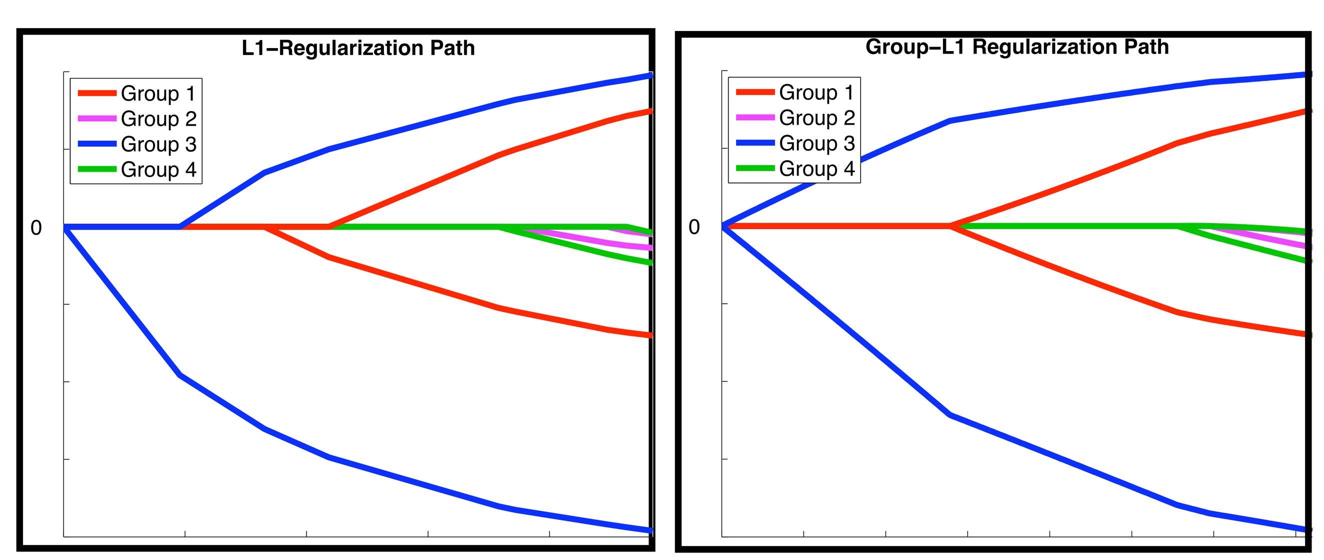
- We use *untied* parameters, and condition on both local and global features

Group-L1 Regularization

- We place an L2 Regularizer on the node parameters and a Group-L1 Regularizer on the edge parameters

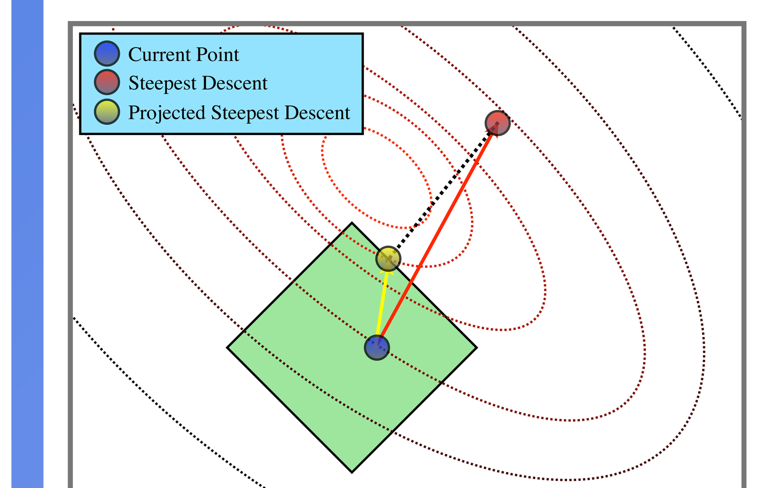
$$\min_{\mathbf{w},\mathbf{v}} -\log p(\mathbf{y}|\mathbf{x}) + \lambda_2 ||\mathbf{w}||_2^2 + \lambda_1 \sum_g ||\mathbf{v}_g||_p$$

- If we use p=2 or p=∞, this leads to group sparsity (edges are removed)



Efficient Optimization

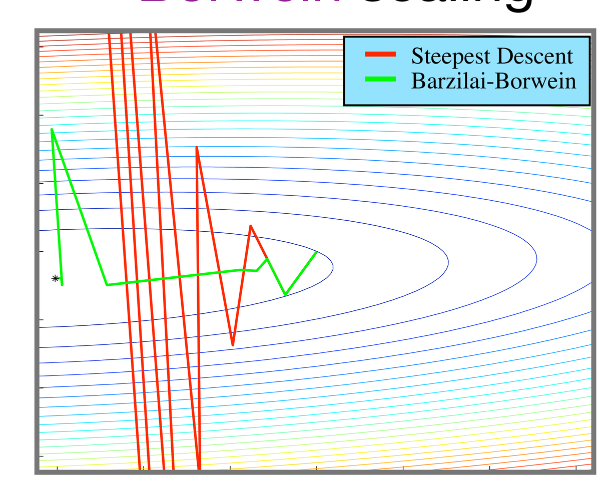
- No existing Group-L1 method satisfies all of the following:
  - (1) they handle a large number of variables
  - (2) they handle a large number of groups
  - (3) they have fast convergence
- We use a novel *Projected Gradient* method that satisfies all 3 of these properties



$$\mathbf{x}_{k+1} := [\mathbf{x}_k - \beta \nabla f(\mathbf{x})]^+$$

$$[\mathbf{x}]^+ := \arg\min_{\mathbf{x}^*} ||\mathbf{x} - \mathbf{x}^*||_2 \quad s.t. \quad \mathbf{x}^* \in C$$

- We use the *Spectral Projected Gradient* method to achieve fast convergence, which uses non-monotone iterations and the Barzilai-Borwein scaling

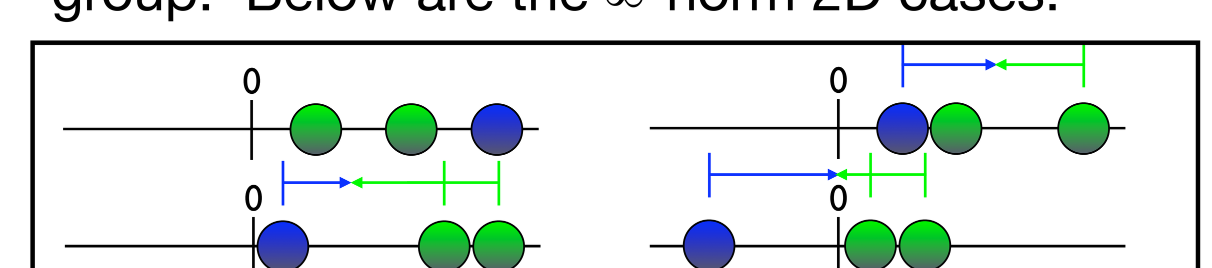


$$\beta = \frac{(\mathbf{x}_k - \mathbf{x}_{k-1})^T(\mathbf{x}_k - \mathbf{x}_{k-1})}{(\mathbf{x}_k - \mathbf{x}_{k-1})^T(\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}))}$$

- We formulate as a constrained optimization by introducing extra variables that bound the p-norms of the individual groups

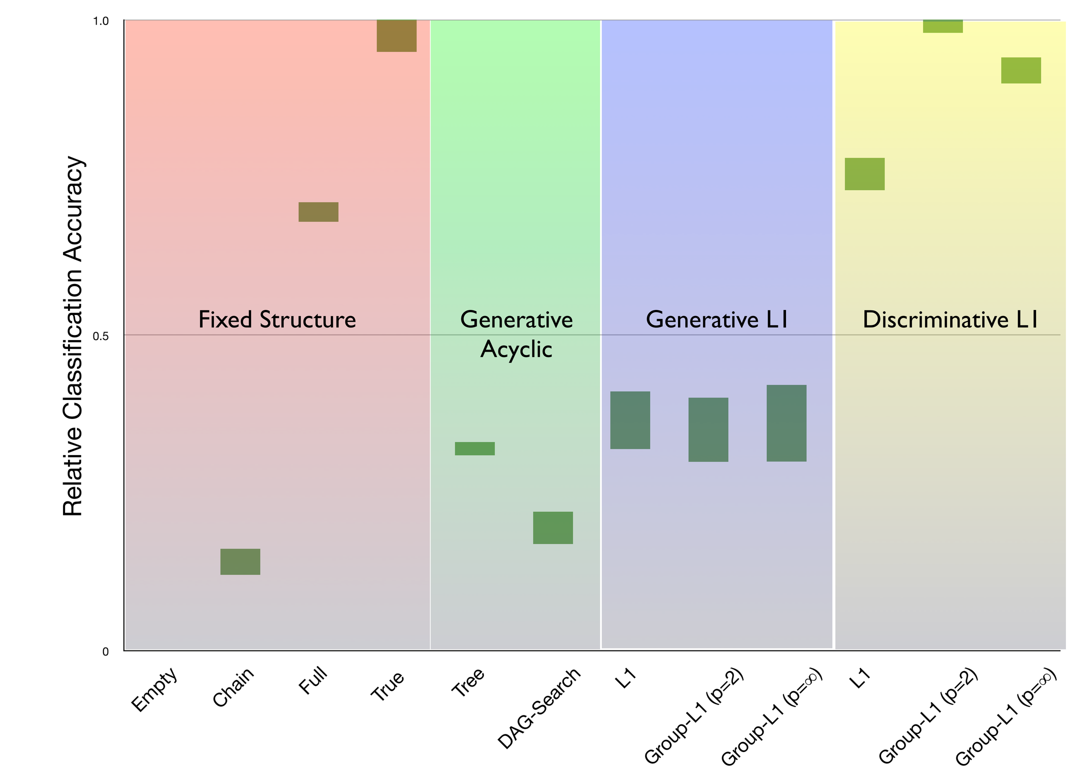
$$\min_{\mathbf{w},\mathbf{v},\alpha} -\log p(\mathbf{y}|\mathbf{x}) + \lambda_2 ||\mathbf{w}||_2^2 + \lambda_1 \sum_g \alpha_g \quad s.t. \quad \forall_g \alpha_g \geq ||\mathbf{v}_g||_p$$

- This formulation lets us handle a large number of variables and groups, since the projection separates into a simple optimization for each group. Below are the ∞-norm 2D cases:

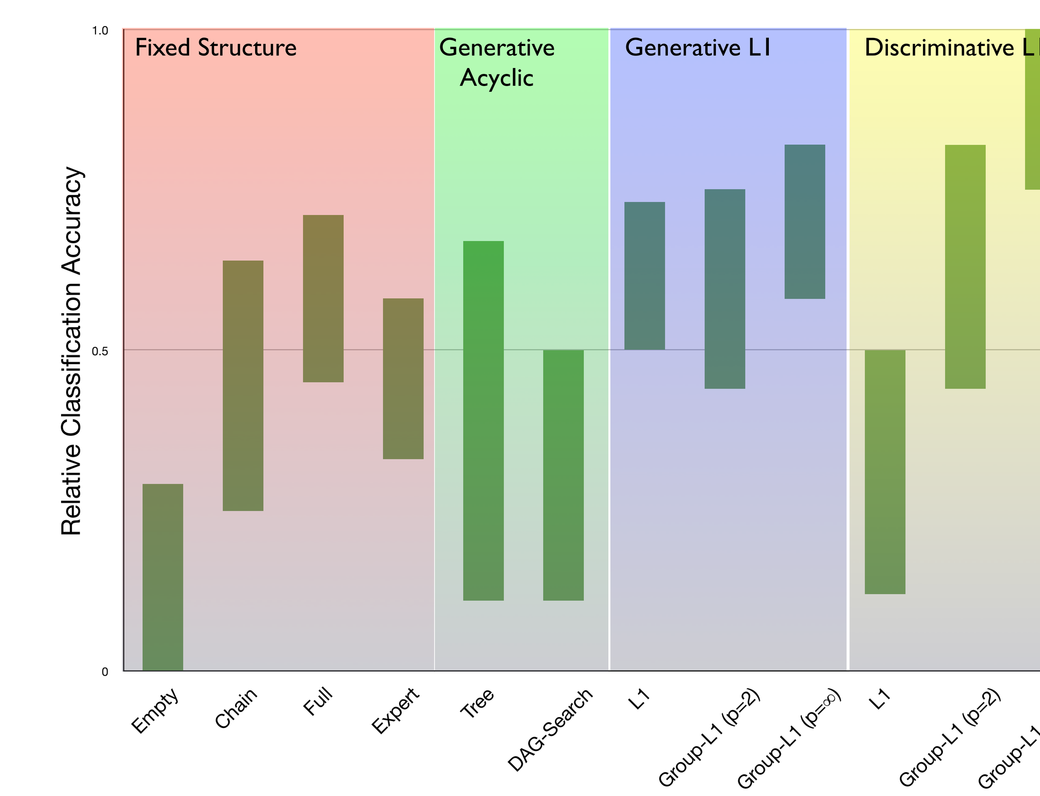


Results

- Comparison of structure learning methods for 10-node synthetic CRF:



- Comparison of structure learning methods for detecting heart motion abnormality:



Conclusion

- Cyclic generative models outperformed acyclic models
- Generative models were not better than the fixed 'Full' structure
- *Discriminatively learned structure with appropriate Group-L1 regularization outperformed generative and fixed structures*