

# On conformational changes of proteins using collective motions in torsion angles and L1 regularization

*Ryabinina R. B.<sup>1</sup>, Emtsev D. I.<sup>2</sup>*

<sup>1</sup> ryabinina.rb@phystech.edu <sup>2</sup> daniil.emcev.ru@yandex.ru

<sup>1,2</sup>MIPT

Investigation of conformational transitions in proteins is an important and well-studied problem in structural bioinformatics with applications ranged from drug design to understanding hidden effects in the structural data. A related important and open question in computational structural bioinformatics is how to efficiently represent transitions between protein structures. Here, we address the problem of how a sparse subset of collective coordinates in the torsion subspace can describe functional conformational changes in proteins. The solution strategy consists in determining the change of torsion angles through the fit of the linearized change of Cartesian coordinates. However, if the fit is not regularized, the structures produced in this approach demonstrate the deviation of several Angstroms from the targets. Rescaled ridge regression (RRR) has been recently introduced to regularize multi-dimensional regressions with correlated explanatory variables. The resulting torsional conformational changes generate conformations that are much more similar to the target conformations. This approach also predicts atomic thermal fluctuations that are better correlated with the ones measured experimentally. Our goal is to find a solution of a ridge regression problem with an L1 regularization constraint using the LASSO formulation. Not much has been done in the torsional angle subspace (internal coordinates) for this problem and nearly nothing has been done using L1 regularization.

## 1 Introduction

There is growing interest in the investigation of the intrinsic dynamic properties of proteins in their native state. They play a key role in ensuring proper functional activity, notably for catalysis, allosteric regulation or molecular recognition. Despite recent progress, the experimental studies of protein dynamics remain rather challenging, and computational methods often constitute valuable alternatives. It is often assumed that torsion angles are the natural degrees of freedom for describing protein motions [1], since bond lengths and bond angles are strongly constrained by covalent forces. Because of this reason, several computational methods have been developed to study protein dynamics in torsion angle space. There is a desperate need of systems to predict dynamic motions. Elastic network models (ENMs) [2–4] are becoming increasingly popular since they provide detailed analytic predictions of native protein dynamics at a very reasonable computational cost.

Ridge regression is one of the most common methods for regularising fits with many variables. It relies heavily on the choice of an adequate value for the ridge (regularization) parameter. Its optimal value is generally unknown, and several criteria have been proposed for its determination. For example, the cross-validation technique is the most popular choice for it.

There are two popular techniques, Tikhonov regularization and ridge regularization which deal with collinearity in multivariate regression. Inspired by successful use of the Lasso method, we propose a LASSO formulation with the direction vectors reconstructed from the internal coordinates. We are using as a base the TNM model [1]. We received better computing using RMSD measure between the prediction and the solution on several benchmarks.

Our aim here is threefold:

1. construct a novel approach with the Lasso regularization;
2. apply this procedure for fitting the B-factors with predicted internal and rigid-body motions, in order to properly calibrate models of protein dynamics and to infer the respective amplitudes of the fluctuations dynamics and to infer the respective amplitudes of the fluctuations;
3. compare results with the previous methods

Compared to the classical variable selection methods, such as subset selection, the LASSO method has two advantages. First, the selection process in the LASSO is based on continuous trajectories of regression coefficients as functions of the penalty level and hence it is more stable than the subset selection methods. Second, the LASSO method is computationally feasible for high-dimensional data [5–7] .

## 2 Methods

### 2.1 Data sets

We use a test set of 30 non-redundant monometric proteins which shows the structural transitions.

### 2.2 Theory part

Theory part

$$Y = X\beta + \varepsilon \quad (1)$$

where  $y \in \mathbb{R}^n$ ,  $\beta \in \mathbb{R}^p$ , and  $X \in \mathbb{R}^{n \times p}$ . We can expand this to  $y_i = \sum_{j=1}^p \beta_j X_{ij} + \varepsilon_i$ ,  $\forall i = 0, 1, \dots, n$ . Here  $\beta_j$  are non-random unknown parameters,  $X_{ij}$  are non-random and observable, and  $\varepsilon_i$  are random so  $y_i$  are random.

From multiple linear regression we have the coefficient estimate

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2)$$

which we can rewrite as  $[(X^T X)^{-1} X^T]^{-1} \hat{\beta} = Y$ .

The LASSO model can be shown in the same form as equation (3) above:

$$(Y - X\beta)^T (Y - X\beta) + \lambda |\beta|_1 \quad (3)$$

Where  $|\beta|_1 = \sum_{j=1}^p |\beta|_j$ .

## References

- [1] R. Mendez and U. Bastolla, “Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins,” *Physical review letters*, vol. 104, p. 228103, Jun 2010.
- [2] M. M. Tirion, “Large amplitude elastic motions in proteins from a single-parameter, atomic analysis,” *Phys. Rev. Lett.*, vol. 77, pp. 1905–1908, Aug 1996.
- [3] A. Atilgan, S. Durell, R. Jernigan, M. Demirel, O. Keskin, and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model,” *Biophysical Journal*, vol. 80, no. 1, pp. 505 – 515, 2001.
- [4] I. Bahar and A. Rader, “Coarse-grained normal mode analysis in structural biology,” *Current Opinion in Structural Biology*, vol. 15, no. 5, pp. 586 – 592, 2005. Carbohydrates and glycoconjugates/Biophysical methods.

- [5] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [6] S. Katayama and S. Imori, “Lasso penalized model selection criteria for high-dimensional multivariate linear regression analysis,” *Journal of Multivariate Analysis*, vol. 132, pp. 138–150, 2014.
- [7] M. R. Osborne, B. Presnell, and B. A. Turlach, “On the lasso and its dual,” tech. rep., Feb. 11 1998.