

A Peaceman-Rachford Splitting Method for the Protein Side-Chain Positioning Problem

Forbes Burkowski* Haesol Im † Henry Wolkowicz †

Sunday 19th March, 2023

Abstract

We formulate a doubly nonnegative (**DNN**) relaxation of the protein side-chain positioning (**SCP**) problem. We inherit the natural splitting of variables that stems from the facial reduction technique in the semidefinite relaxation. We solve the relaxation using a variant of the Peaceman-Rachford splitting method. Our numerical experiments show that we solve *almost all* instances of the NP-hard **SCP** problem to *optimality*.

Keywords: Protein structure prediction, Side-chain positioning, Doubly-nonnegative relaxation, facial reduction, Peaceman-Rachford splitting method

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 1.1 | Notation | 3 |
| 1.2 | Contributions and Outline | 3 |
| 2 | Model Derivation | 4 |
| 2.1 | Problem Formulation as IQP | 4 |
| 2.2 | SDP Relaxation | 5 |
| 2.2.1 | Gangster Constraint $G_{\hat{J}}(Y) = E_{00}$ | 6 |
| 2.2.2 | Facial Reduction | 7 |
| 2.3 | DNN Relaxation | 7 |
| 3 | The Algorithm | 10 |
| 3.1 | Update Formulae | 10 |
| 3.1.1 | R -Update | 11 |
| 3.1.2 | Y -Update | 11 |
| 3.2 | Bounding | 11 |
| 3.2.1 | Lower Bounds from Lagrange Relaxation | 11 |
| 3.2.2 | Upper Bounds from Nearest Binary Feasible Solutions | 12 |
| 4 | Numerical Experiments with Real-World Data | 12 |
| 4.1 | Stopping Criteria and Parameter Settings | 13 |
| 4.2 | Energy Matrix Computation | 13 |
| 4.2.1 | Removing Collisions | 14 |
| 4.3 | Experiments with Real-World Data | 15 |
| 4.3.1 | A Tighter Relaxation | 16 |

*Professor Emeritus, Cheriton School of Computer Science, University of Waterloo, Canada

†Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Canada, Research supported by NSERC

| | | |
|----------|----------------------------|-----------|
| 5 | Conclusions | 17 |
| | Index | 18 |
| | Bibliography | 19 |
| A | Additional Numerics | 21 |

List of Figures

| | | |
|-----|---|---|
| 2.1 | A Diagram of the Protein Side-Chain Positioning Problem | 4 |
| 2.2 | An illustration of the index set \mathcal{J} ; members of \mathcal{J} correspond to the off-diagonal elements of diagonal blocks of W indicated by the symbol \circ | 6 |

List of Tables

| | | |
|-----|--|----|
| 4.1 | Computational results on selected PDB instances | 16 |
| 4.2 | The solver optimal values of the DNN and SDP relaxations on selected instances | 16 |
| A.1 | Computation results on selected PDB instances up to 100 amino acids | 21 |
| A.2 | Computation results on selected PDB instances up to 200 amino acids | 22 |
| A.3 | Computation results on selected PDB instances up to 300 amino acids | 23 |

1 Introduction

The *protein side-chain positioning* (**SCP**) problem is one of the most important subproblems of the protein structure prediction problem. We formulate the **SCP** problem as an integer program and derive its doubly nonnegative, **DNN**, relaxation. We then use a variation of the *Peaceman-Rachford splitting method* (**PRSM**) to solve the **DNN** relaxation.

The applications of **SCP** extend to ligand binding [18, 21] and protein-protein docking with backbone flexibility [22, 27]. A protein is a macromolecule consisting of a long main chain backbone that provides a set of anchors for a sequence of amino acid side-chains. The backbone is comprised of a repeating triplet of atoms (nitrogen, carbon, carbon) with the central carbon atom being designated as the alpha carbon. An amino acid side-chain is a smaller (1 to 18 atoms) side branch that is anchored to an alpha carbon. The positions of the atoms in a side-chain can be established by knowing the 3D position of its alpha carbon and the dihedral angles defined by atoms in the side-chain. The number of dihedral angles varies from 1 to 4 depending on the length of the side-chain. This is true for 18 of the 20 amino acids with glycine and alanine being exceptions because their low atom counts preclude dihedral angles.

It has been observed that the values of dihedral angles are not uniformly distributed. They tend to form clusters with cluster centers that are equally separated (+60, 180, -60). Consequently, if the dihedral angles are unknown we at least have a reasonable estimate of their values by appealing to these discretized values. With this strategy being applied, a side-chain with one dihedral angle would have three possible sets of positions for its atoms. We refer to each set of atomic positions as a rotamer. A side-chain with two dihedral angles will have 3 times 3 or 9 different arrangements of the atoms (i.e. 9 rotamers). Three dihedral angles will result in 27 rotamers and four dihedral angles will give 81 rotamers.

In the **SCP** problem we are given a fixed backbone and a designation of the amino acid type for each alpha carbon. To solve the problem it is required that each amino acid is assigned a particular rotameric setting with the objective of avoiding any collisions with neighbouring amino

acids that are given their rotameric settings. Avoiding collisions will lower the overall energy of the protein and, in fact, even with all possible collisions circumvented we want to have an energy evaluation that is minimal.

The **SCP** problem is proven to be NP-hard [1]. The nature of the **SCP** problem has motivated the development of many heuristic based algorithms [3,4,7,9,24,28] and many of these approaches rely on the graph structure of the problem. Other approaches for solving **SCP** problems have been proposed. These range from probabilistic approaches [16, 19, 25], integer programming [2, 13, 17], to semidefinite programming [5,8]. Our approach is based on a semidefinite programming relaxation. Given a rotamer library, the **SCP** problem can be formulated as an *integer quadratic problem (IQP)*. We then obtain a *semidefinite programming (SDP)* relaxation to the **IQP** via a lifting of variables and *facial reduction (FR)*. We finally obtain a *doubly nonnegative (DNN)* relaxation by adding nonnegativity constraints and some additional constraints to the **SDP** that help strengthen our relaxation.

The facial reduction originating from the **SDP** relaxation delivers a natural splitting of variables. This elegant splitting of variables fits into the framework of the splitting methods. The framework gives an efficient procedure of engaging constraints that are difficult to process simultaneously, see e.g., [15, 20, 23]. We solve the **DNN** relaxation using a variation of the so-called *Peaceman-Rachford splitting method (PRSM)*. Using the **PRSM**, we examine the strength of our approach in the numerical experiments. The usage of splitting method for the **DNN** relaxation allows for an effective treatment for handling implicit redundant constraints and the ill-posed data that stems from collisions between rotamers.

1.1 Notation

We let $\mathbb{R}^n, \mathbb{R}^{m \times n}$ denote the standard real Euclidean spaces; \mathbb{S}^n denotes the Euclidean space of n -by- n real symmetric matrices; \mathbb{S}_+^n (\mathbb{S}_{++}^n , resp) denotes the cone of n -by- n positive semidefinite (definite, resp) matrices. We write $X \succeq 0$ if $X \in \mathbb{S}_+^n$, and $X \succ 0$ if $X \in \mathbb{S}_{++}^n$. We use $\text{range}(X)$ and $\text{null}(X)$ to denote the range of X and the null space of X , respectively. Given $X \in \mathbb{R}^{n \times n}$, we use $\text{trace}(X)$ to denote the *trace* of X . Given two matrices $X, Y \in \mathbb{R}^{m \times n}$, we let $\langle X, Y \rangle = \text{trace}(XY^T)$ denote the usual trace inner product between X and Y ; $X \circ Y$ denote the element-wise, or Hadamard, product of X and Y . Given a closed convex set \mathcal{C} in a Euclidean space, we let $\mathcal{N}_{\mathcal{C}}(x)$ denote the *normal cone* at $x \in \mathcal{C}$ with respect to \mathcal{C} . Given $X \in \mathbb{R}^{n \times n}$, we use $\text{diag}(X)$ to denote the vector formed from the diagonal entries of X . Then $\text{Diag}(v) = \text{diag}^*(v)$ is the adjoint linear transformation that forms the diagonal matrix from the vector v . Given a collection of matrices $\{A_i\}_{i=1}^m$, we let $\text{BlkDiag}(A_1, \dots, A_m)$ denote the block diagonal matrix with the i -th diagonal block A_i . We let \bar{e}_n denote the n -dimensional vector with each entry set to 1 and we omit the subscript when the dimension is clear. Given a positive integer m , we often use the notation $[m]$ to mean the set of positive integers $\{1, \dots, m\}$.

1.2 Contributions and Outline

We present the process for formulating the model in Section 2. We formulate the **SCP** problem as an integer quadratic program, **IQP**, and obtain the **SDP** and **DNN** relaxations. The derivation for the **SDP** relaxation is first presented in [5] via Lagrangian relaxation. Here, we present a much simpler derivation for the **SDP** relaxation via direct lifting of the variables. In Section 3 we present a variation of the **PRSM** for solving the **DNN** relaxation as well as our strategies to obtain upper and lower bounds to the **SCP** problem. We show that the splitting method engages implicit redundant constraints safely that arise from the facial reduction. In Section 4 we use the

real-world data from the Protein Data Bank¹ to examine the strength of our approach. We show that the usage of splitting method to the **DNN** relaxation effectively handles collisions between rotamers that are indicated by large values in the data. Moreover, the numerical experiments demonstrate that our approach *provably* solves *almost all* instances² to the global optimum of the NP-hard protein **SCP** problem.

2 Model Derivation

The goal of this section is to obtain the **DNN** relaxation of the **SCP** problem. We start by presenting a formulation of the **SCP** as an **IQP** in Section 2.1. We then derive its **SDP** relaxation in Section 2.2. We continue the derivation by identifying redundant constraints in the **IQP** and in the **SDP** relaxation in order to obtain a complete (stable) **DNN** relaxation in Section 2.3.

2.1 Problem Formulation as IQP

We are given a collection of disjoint sets $\mathcal{V}_i, i = 1, \dots, p$. Each set \mathcal{V}_i has m_i members and we index its members

$$\mathcal{V}_i := \{v_i^1, v_i^2, \dots, v_i^{m_i}\}, \text{ for all } i = 1, \dots, p.$$

We call each set \mathcal{V}_i a *rotamer set* and its members *rotamers*. We use $n_0 = \sum_{i=1}^p m_i$ and $\mathcal{V} = \cup_{i=1}^p \mathcal{V}_i$. The *protein side-chain positioning problem* seeks to

1. select *exactly one* rotamer v_i^j , from each set \mathcal{V}_i , where $j \in [m_i]$ (see Figure 2.1³); and
2. minimize the sum of the weights (energy) determined by chosen rotamers, and the energy between each chosen rotamer and the backbone.

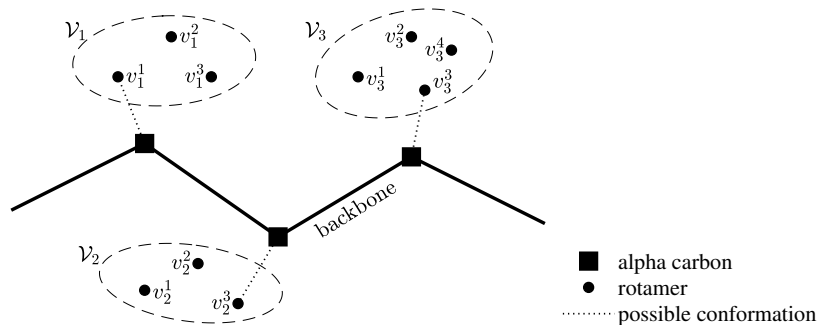


Figure 2.1: A Diagram of the Protein Side-Chain Positioning Problem

Viewing the rotamers as a set of nodes of a graph, we can realize the **SCP** problem as a discrete optimization problem over a graph. We construct a matrix $E \in \mathbb{S}^{n_0}$ to record the energy values between rotamers and the backbone. We use the matrix entries E_{uv} , with $u \neq v$, to denote the edge weights between two distinct rotamers (nodes); while the diagonal entries E_{uu} denote the weight between the rotamer u and the backbone. This yields a symmetric matrix E , where $E_{uv} = \infty$ if both rotamers u, v are in the same set. We note that the multiplication $0 \cdot \infty = 0$ when adding up the weights (energies). Alternatively, we can set these weights to 0 and add a

¹<https://www.rcsb.org/>

²Out of 131 test problems, one problem had a positive gap; five other problems had gaps of approximately 10^{-6} .

³ \mathcal{V}_i indicates the i -th rotamer set and v_i^j indicates the j -th candidate in the i -th rotamer set \mathcal{V}_i .

constraint to choose exactly one rotamer from each set, which is what we do. Thus each diagonal block of E , of size m_i , can be assumed to be a diagonal matrix. We can make this simplification without loss of generality since we are looking to only choose one rotamer per set \mathcal{V}_i .

We are looking to solve the following integer quadratic program over the indicator vector x :

$$\begin{aligned}
p_{\mathbf{IQP}}^* &:= \min_x \sum_{u,v} E_{uv} x_u x_v \\
\text{s.t.} \quad &\sum_{u \in \mathcal{V}_k} x_u = 1, \quad k = 1, \dots, p \\
&x = [w_1^T \quad w_2^T \quad \dots \quad w_p^T]^T \\
&w_i \in \{0, 1\}^{m_i}, \quad i = 1, \dots, p.
\end{aligned} \tag{2.1}$$

The constraints in (2.1) forces that exactly one element of v_i is set to be 1, consequently modelling that exactly one rotamer is chosen for each rotamer set \mathcal{V}_i . We construct the block diagonal matrix

$$A = \text{BlkDiag}(\bar{e}_{m_1}^T, \bar{e}_{m_2}^T, \dots, \bar{e}_{m_p}^T) \in \mathbb{R}^{p \times n_0}. \tag{2.2}$$

We then use $Ax = b$ to work with a concise representation of the first equality constraint in (2.1). Finally, we obtain the following representation of the **SCP** problem:

$$\begin{aligned}
(IQP) \quad p_{\mathbf{IQP}}^* &= \min_x x^T E x \\
&\text{s.t.} \quad Ax = \bar{e}_p \\
&\quad x \in \{0, 1\}^{n_0}.
\end{aligned} \tag{2.3}$$

2.2 SDP Relaxation

The problem (2.3) is NP-hard and hence we resort to a relaxation. We define

$$\hat{E} := \text{BlkDiag}(0, E) \in \mathbb{S}^{n_0+1}, \quad E_{00} := e_0 e_0^T \in \mathbb{S}^{n_0+1},$$

where e_0 is the first unit vector. In this section we aim to obtain the following **SDP** relaxation to the discrete optimization problem (2.3):

$$\begin{aligned}
(SDP) \quad p_{\mathbf{SDP}}^* &:= \min_{R, Y} \text{trace}(\hat{E}Y) \\
&G_{\hat{J}}(Y) = E_{00} \\
&Y = VRV^T \\
&R \in \mathbb{S}_+^{n_0+1-p},
\end{aligned} \tag{2.4}$$

where $G_{\hat{J}}(\cdot)$ and V are explained in Section 2.2.1 and Section 2.2.2, respectively. A variant of the relaxation (2.4) is proposed by [5] via Lagrangian relaxation to (2.3). Here we present a simpler derivation of the model (2.4) via a simple direct lifting.

The first step for deriving the **SDP** relaxation (2.4) is to lift the variable dimension. Given $x \in \mathbb{R}^{n_0}$, we lift to symmetric matrix space using the rank-one *lifted matrix*

$$Y_x := \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T = \begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} \in \mathbb{S}^{n_0+1}.$$

For the **SDP** relaxation, we index the rows and columns *starting from 0*, i.e., the row and column indices are $\{0, 1, \dots, n_0\}$. This lifting allows for an alternative representation of the objective function

$$x^T E x = \left\langle \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix}, \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T \right\rangle = \langle \hat{E}, Y_x \rangle.$$

In the remaining of this section we show how this lifting process gives rise to the constraints of the model (2.4):

1. the linear (gangster) constraint $G_{\hat{\mathcal{J}}}(Y) = E_{00}$ (Section 2.2.1);
2. $Y = VRV^T$, where $R \in \mathbb{S}_+^{n_0+1-p}$ (Section 2.2.2).

2.2.1 Gangster Constraint $G_{\hat{\mathcal{J}}}(Y) = E_{00}$

Given a matrix $W \in \mathbb{S}^{n_0}$, we define the set of indices

$$\mathcal{J} := \left\{ \left(\sum_{i=1}^j m_{i-1} + k, \sum_{i=1}^j m_{i-1} + \ell \right) : j \in \{1, \dots, p-1\}, k, \ell \in \{2, \dots, m_j - 1\}, k \neq \ell \right\}.$$

Here, m_i is the cardinality of rotamer set \mathcal{V}_i , and $m_0 = 0$. In other words, \mathcal{J} is the set of off-diagonal indices of the m_i -by- m_i diagonal blocks of $W \in \mathbb{S}^{n_0}$; see Figure 2.2 for a visual

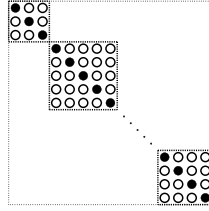


Figure 2.2: An illustration of the index set \mathcal{J} of 0's; members of \mathcal{J} correspond to the off-diagonal elements of diagonal blocks of W indicated by the symbol \circ .

illustration of the positioning of these indices. Note that these indices correspond to exactly

$$W_{uv} = x_u x_v = 0, u \neq v, u, v \in \mathcal{V}_i,$$

i.e., the constraint on any two distinct rotamers in the same rotamer set cannot be chosen.

With the above set of indices, we define the mapping

$$G_{\mathcal{J}} : \mathbb{S}^{n_0} \rightarrow \mathbb{R}^{|\mathcal{J}|} \quad \text{by } G_{\mathcal{J}}(W) = (W_{ij})_{ij \in \mathcal{J}}.$$

By abuse of notation, we also view the mapping $G_{\mathcal{J}}$ as an operator from \mathbb{S}^{n_0} to \mathbb{S}^{n_0} to mean

$$G_{\mathcal{J}} : \mathbb{S}^{n_0} \rightarrow \mathbb{S}^{n_0}, \quad (G_{\mathcal{J}}(W))_{i,j} = \begin{cases} W_{i,j} & \text{if } (i,j) \text{ or } (j,i) \in \mathcal{J}, \\ 0 & \text{otherwise.} \end{cases}$$

The map $G_{\mathcal{J}}$ can also be viewed as the operator on \mathbb{S}^{n_0} defined by $G_{\mathcal{J}}(W) = (A^T A - I) \circ W$ with A defined in (2.2). Recall \circ is the element-wise matrix product. In plain words, $G_{\mathcal{J}}(W)$ is the projection that chooses elements of W corresponding to the index set \mathcal{J} . The constraint $G_{\mathcal{J}}(W) = 0$ is often called the *gangster constraint* and it is due to the fact that elements of W associated with \mathcal{J} are set to be zero (shoots holes in the matrix).

We now define the set of pairs of indices

$$\hat{\mathcal{J}} := \{(0,0)\} \cup \mathcal{J} \subset \{0, 1, \dots, n_0\} \times \{0, 1, \dots, n_0\}$$

to directly work with the lifted variable in \mathbb{S}^{n_0+1} . We define the analogous mapping $G_{\hat{\mathcal{J}}}$ with $\hat{\mathcal{J}}$:

$$G_{\hat{\mathcal{J}}} : \mathbb{S}^{n_0+1} \rightarrow \mathbb{R}^{|\hat{\mathcal{J}}|} \quad \text{by } G_{\hat{\mathcal{J}}}(Y) = (Y_{ij})_{ij \in \hat{\mathcal{J}}}.$$

This yields the *gangster constraint* in projection and operator equivalent forms, respectively,

$$G_{\hat{\mathcal{J}}}(Y) = e_0 \in \mathbb{R}^{1+|\hat{\mathcal{J}}|}, \quad G_{\hat{\mathcal{J}}}(Y) = E_{00}.$$

2.2.2 Facial Reduction

We now derive the constraint $Y = VRV^T$ and $R \in \mathbb{S}_+^{n_0+1-p}$. Let x be a feasible solution to (2.3) and we observe the following implications:

$$\begin{aligned} Ax = \bar{e}_p &\implies \begin{bmatrix} 1 \\ x \end{bmatrix}^T \begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix} = 0 \\ &\implies \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T \begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix} \begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix}^T = 0 \\ &\implies \left\langle \underbrace{\begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T}_{=: Y_x}, \underbrace{\begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix} \begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix}^T}_{=: K} \right\rangle = 0. \end{aligned}$$

Since both arguments in the last inner product are positive semidefinite, we obtain the useful property:

$$\langle K, Y_x \rangle = 0 \implies KY_x = 0 \implies \text{range}(Y_x) \subseteq \text{null}(K). \quad (2.5)$$

In other words, $\text{null}(K)$ captures the range that the feasible points can have.

We now exploit the property (2.5) to restrict the range of the variable. We find a full-column rank matrix $V \in \mathbb{R}^{(n_0+1) \times (n_0+1-p)}$ such that

$$\text{range}(V) = \text{null}(K) = \text{null} \left(\begin{bmatrix} -\bar{e}_p^T \\ A^T \end{bmatrix}^T \right).$$

For our purposes we choose V with normalized columns. Since A is full-row rank, we get that $\text{rank}(K) = p$. Finally, we can capture any feasible Y_x using V :

$$Y_x \in V\mathbb{S}_+^{n_0+1-p}V^T.$$

This is the well-known *facial reduction* technique, see e.g., [10]. The matrix K functions as an *exposing vector* for the feasible set. The matrix V is known as a facial range vector.

The remaining step for the **SDP** relaxation is simple. We note that $\text{rank}(Y_x) = 1$ and this leaves the feasible region nonconvex. We discard the rank restriction on the variable Y_x to work with a convex feasible region and the variable of the form

$$Y = VRV^T \text{ where } R \in \mathbb{S}_+^{n_0+1-p}.$$

This completes the derivation of the relaxation in (2.4). It is known that there is a $\hat{R} \in \mathbb{S}_{++}^{n_0+1-p}$ feasible to (2.4); see [5].

2.3 DNN Relaxation

We continue with the **SDP** relaxation derived in Section 2.2 to complete our relaxation by adding additional constraints to (2.4). In Theorem 2.1 below, we obtain two additional properties of the model (2.4).

Theorem 2.1. *Suppose that (R, Y) are feasible to (2.4). Then the following hold.*

1. *The first column of Y is equal to the diagonal of Y .*
2. $\text{trace}(R) = 1 + p$.

Proof. We recall that $\text{range}(V) = \text{null}(K) = \text{null}([- \bar{e}_p \ A])$. Hence we have

$$[- \bar{e}_p \ A] Y = [- \bar{e}_p \ A] V R V^T = 0 R V^T = 0. \quad (2.6)$$

We then exploit the structure of $[- \bar{e}_p \ A] Y$. We first partition Y as follows:

$$Y = \begin{bmatrix} 1 & Y_{10}^T & Y_{20}^T & \cdots & Y_{p0}^T \\ Y_{10} & Y_{11} & Y_{12} & \cdots & Y_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{p0} & Y_{p1} & Y_{p2} & \cdots & Y_{pp} \end{bmatrix} \in \mathbb{S}^{n_0+1}, \quad (2.7)$$

where $Y_{ii} \in \mathbb{S}^{m_i}$, $Y_{ij} \in \mathbb{R}^{m_i \times m_j}$, $Y_{i0} \in \mathbb{R}^{m_i}$, $\forall i, j \in [p]$. We use $Y_{ij}^{\text{col } \ell}$ to denote the ℓ -th column of the (i, j) -th block of Y and $Y_{i0, \ell}$ to denote the ℓ -th coordinate of the vector $Y_{i0} \in \mathbb{R}^{m_i}$. Then expanding $[- \bar{e}_p \ A] Y$ with the block representation (2.7) yields

$$[- \bar{e}_p \ A] Y = [a_0 \ A_1 \ \cdots \ A_p] \in \mathbb{R}^{p \times (n_0+1)},$$

where

$$a_0 = \begin{bmatrix} -1 + \bar{e}_{m_1}^T Y_{10} \\ -1 + \bar{e}_{m_1}^T Y_{20} \\ \vdots \\ -1 + \bar{e}_{m_p}^T Y_{p0} \end{bmatrix} \in \mathbb{R}^p, \quad (2.8)$$

and, for each $i \in [p]$,

$$A_i = \begin{bmatrix} -Y_{i0,1} + \bar{e}_{m_1}^T Y_{1i}^{\text{col } 1} & -Y_{i0,2} + \bar{e}_{m_1}^T Y_{1i}^{\text{col } 2} & \cdots & -Y_{i0,m_i} + \bar{e}_{m_1}^T Y_{1i}^{\text{col } m_i} \\ \vdots & \vdots & \ddots & \vdots \\ -Y_{i0,1} + \bar{e}_{m_j}^T Y_{ji}^{\text{col } 1} & -Y_{i0,2} + \bar{e}_{m_j}^T Y_{ji}^{\text{col } 2} & \cdots & -Y_{i0,m_i} + \bar{e}_{m_j}^T Y_{ji}^{\text{col } m_i} \\ \vdots & \vdots & \ddots & \vdots \\ -Y_{i0,1} + \bar{e}_{m_p}^T Y_{1p}^{\text{col } 1} & -Y_{i0,2} + \bar{e}_{m_p}^T Y_{1p}^{\text{col } 2} & \cdots & -Y_{i0,m_i} + \bar{e}_{m_p}^T Y_{1p}^{\text{col } m_i} \end{bmatrix} \in \mathbb{R}^{p \times m_i}.$$

By (2.6), we have $A_i = 0$, $\forall i \in [p]$. Thus, for each $i \in [p]$, the i -th row of A_i yields

$$Y_{i0, \ell} = \bar{e}_{m_i}^T Y_{ii}^{\text{col } \ell}, \quad \ell \in [m_i].$$

Since $G_{\hat{J}}(Y) = E_{00}$ holds, we see that

$$\text{diag}(Y_{ii}) = Y_{i0}, \quad \forall i \in [p].$$

Therefore, we conclude that the first column and the diagonal of Y are identical.

We now show that $\text{trace}(R) = 1 + p$. By (2.6), we have the vector a_0 from (2.8) is 0. Thus, $1 = \bar{e}_{m_i}^T$, for all $i = 1 \dots, p$. By Item 1, we obtain

$$\bar{e}_{m_i}^T Y_{i0} = 1, \quad \forall i \in [p].$$

Since $\text{diag}(Y_{ii}) = Y_{i0}$, $\forall i \in [p]$, we must have that $\text{trace}(Y_{ii}) = 1$, $\forall i \in [p]$. Hence $Y = V R V^T$ gives

$$1 + p = \text{trace}(Y) = \text{trace}(V R V^T) = \text{trace}(R),$$

where the last equality holds since $V^T V = I$. \square

Item 1 of Theorem 2.1 is known in the literature; see [5]. This property is discovered by using the Lagrangian dual. Here, we displayed an alternative derivation that exploits the steps during the direct lifting.

We recall that the original model (2.3) has the binary constraint on its variable x . We also recall that the direct lifting yields a variable of the form $\begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} \in \mathbb{S}^{n_0+1}$. Hence, we strengthen our model by including the constraint $Y_{i,j} \in [0, 1], \forall i, j$.

We define the sets

$$\begin{aligned} \mathcal{R} &:= \{R \in \mathbb{S}^{n_0+1-p} : R \succeq 0, \text{trace}(R) = p + 1\}, \\ \mathcal{Y} &:= \{Y \in \mathbb{S}^{n_0+1} : G_{\hat{j}}(Y) = E_{00}, 0 \leq Y \leq 1\}. \end{aligned}$$

By including additional constraints $\text{trace}(R) = 1 + p$ and $0 \leq Y \leq 1$ to the **SDP** relaxation (2.4), we complete our model, the **DNN** relaxation to (2.3):

$$\begin{aligned} (DNN) \quad p_{\text{DNN}}^* &:= \min_{R, Y} \text{trace}(\hat{E}Y) \\ &Y = VRV^T \\ &R \in \mathcal{R} \\ &Y \in \mathcal{Y}. \end{aligned} \tag{2.9}$$

We remark that both **(DNN)** and **(SDP)** are relaxations to **(IQP)**, but **(DNN)** is a strengthened model than **(SDP)**, i.e.,

$$p_{\text{SDP}}^* \leq p_{\text{DNN}}^* \leq p_{\text{IQP}}^*.$$

We also remark that there are redundant constraints in the model (2.9). These (implicit) redundant constraints result in numerical instabilities when they are not treated carefully. In Section 2.3 below, we use the splitting method to distribute constraints to two different subproblems. We benefit from the usage of the splitting method in two distinct ways; we handle the numerically difficult problem into two separate easier subproblems; and we avoid the numerical instabilities that arise from the redundant constraints.

We demonstrate the strength of **(DNN)** in Section 4.3.1. The **DNN** relaxation has a linear objective with an *onto* linear equality constraint, and compact, convex, feasible set constraints. The first-order optimality conditions for (2.9) are

$$\begin{aligned} 0 &\in -V^T Z V + \mathcal{N}_{\mathcal{R}}(R), && \text{(dual feasibility with respect to } R) \\ 0 &\in \hat{E} + Z + \mathcal{N}_{\mathcal{Y}}(Y), && \text{(dual feasibility with respect to } Y) \\ Y &= \hat{V} R \hat{V}^T, \quad R \in \mathcal{R}, Y \in \mathcal{Y}, && \text{(primal feasibility)} \end{aligned} \tag{2.10}$$

where $\mathcal{N}_{\mathcal{R}}(R), \mathcal{N}_{\mathcal{Y}}(Y)$ are the normal cones and Z is a Lagrange multiplier associated with the constraint $Y = VRV^T$. Theorem 2.2 below states that some elements of the dual optimal multiplier Z^* are known in advance.

Theorem 2.2. *Let (R^*, Y^*) be an optimal pair for (2.9), and let*

$$\mathcal{Z}_A := \left\{ Z \in \mathbb{S}^{n_0+1} : Z_{i,i} = -(\hat{E})_{i,i}, Z_{0,i} = Z_{i,0} = -(\hat{E})_{0,i}, i = 1, \dots, n_0 \right\}.$$

Then there exists $Z^ \in \mathcal{Z}_A$ such that (R^*, Y^*, Z^*) solves (2.10).*

Proof. The proof uses the optimality conditions (2.10) and Theorem 2.1. The proof can be found in [15, Theorem 2.11]. \square

3 The Algorithm

In this section we present the algorithm for solving the **DNN** relaxation (2.9). For $\beta > 0$, we define the augmented Lagrangian \mathcal{L}_A of the model (2.9):

$$\mathcal{L}_A(R, Y, Z) := \langle \hat{E}, Y \rangle + \langle Z, Y - VRV^T \rangle + \frac{\beta}{2} \|Y - VRV^T\|_F^2. \quad (3.1)$$

We define the projection operator $\mathcal{P}_{\mathcal{Z}_0}(Z)$ onto the set

$$\mathcal{Z}_0 = \{Z \in \mathbb{S}^{n_0+1} : Z_{i,i} = Z_{0,i} = Z_{i,0} = 0, i = 1, \dots, n_0\}.$$

In other words, the projection operator $\mathcal{P}_{\mathcal{Z}_0}(Z)$ sets the first column, first row and the diagonal elements of Z to be 0, except for the (0,0)-th entry.

We use the *restricted dual Peaceman-Rachford splitting method*, **rPRSM** (Algorithm 3.1), a variation of the strictly contractive Peaceman-Rachford splitting method to solve the model (2.9). We note that the ordinary **PRSM** updates the dual multipliers without the projection operator

Algorithm 3.1 **rPRSM** [15] for solving (2.9)

Initialize: $Y^0 \in \mathbb{S}^{n_0+1}$, $Z^0 \in \mathcal{Z}_A$, $\beta \in (0, \infty)$, $\gamma \in (0, 1)$

while termination criteria are not met **do**

$$R^{k+1} = \underset{R \in \mathcal{R}}{\operatorname{argmin}} \mathcal{L}_A(R, Y^k, Z^k)$$

$$Z^{k+\frac{1}{2}} = Z^k + \gamma\beta \cdot \mathcal{P}_{\mathcal{Z}_0}(Y^k - VR^{k+1}V^T)$$

$$Y^{k+1} = \underset{Y \in \mathcal{Y}}{\operatorname{argmin}} \mathcal{L}_A(R^{k+1}, Y, Z^{k+\frac{1}{2}})$$

$$Z^{k+1} = Z^{k+\frac{1}{2}} + \gamma\beta \cdot \mathcal{P}_{\mathcal{Z}_0}(Y^{k+1} - VR^{k+1}V^T)$$

end while

$\mathcal{P}_{\mathcal{Z}_0}$. The projection on the dual multiplier Z is motivated from an endeavour to have a better dual multiplier at each iteration. We recall that some of the elements of the optimal dual multipliers are known by Theorem 2.2. The algorithm fixes these known elements to be the optimal elements at every iteration. We leave the details of the convergence proof of **rPRSM** scheme to [15, Theorem 3.2].

Remark 3.1. *The model (2.9) can be solved by using a standard **SDP** solver. The nonnegativity of each element of Y is considered using cutting planes in [5]. However, this approach becomes more computationally challenging as the number of cutting planes increases. Splitting methods engage the polyhedral constraints $0 \leq Y \leq 1$ in an economic manner. We incorporate the positive semidefinite constraint and the nonnegativity constraint very efficiently. We deal with the positive semidefinite and trace constraint in the R -subproblem, and then deal with the interval and gangster constraints in the Y -subproblem.*

3.1 Update Formulae

In this section we present the formulae for the R and Y updates in Algorithm 3.1. The update rules are discussed in [15]. We include the formulae for completeness.

3.1.1 R-Update

In this section we present the update rule for the R -subproblem. The formula for the R -subproblem, with \mathcal{L}_A defined in (3.1), is as follows:

$$\begin{aligned}
R^{k+1} &= \operatorname{argmin}_{R \in \mathcal{R}} \mathcal{L}_A(R, Y^k, Z^k) \\
&= \operatorname{argmin}_{R \in \mathcal{R}} \left\| Y^k - V R V^T + \frac{1}{\beta} Z^k \right\|_F^2 \\
&= \operatorname{argmin}_{R \in \mathcal{R}} \left\| R - V^T \left(Y^k + \frac{1}{\beta} Z^k \right) V \right\|_F^2 \\
&= \mathcal{P}_{\mathcal{R}} \left(V^T \left(Y^k + \frac{1}{\beta} Z^k \right) V \right) \\
&= U \operatorname{Diag} \left(\mathcal{P}_{\Delta_{p+1}}(d) \right) U^T,
\end{aligned}$$

where the second equality holds by completing the square; the third equality holds due to $V^T V = I$; and the last equality follows from the eigenvalue decomposition

$$V^T \left(Y^k + \frac{1}{\beta} Z^k \right) V = U \operatorname{Diag}(d) U^T,$$

and $\mathcal{P}_{\Delta_{p+1}}(\cdot)$ is the projection operator onto the simplex $\Delta_{p+1} = \{z \in \mathbb{R}^{n_0+1-p} : \bar{e}^T z = 1 + p\}$.

3.1.2 Y-Update

The update rule for Y is as follows:

$$\begin{aligned}
Y^{k+1} &= \operatorname{argmin}_{Y \in \mathcal{Y}} \mathcal{L}_A(R^{k+1}, Y, Z^{k+\frac{1}{2}}) \\
&= \operatorname{argmin}_{Y \in \mathcal{Y}} \left\| Y - \left(V R^{k+1} V^T - \frac{1}{\beta} (\hat{E} + Z^{k+\frac{1}{2}}) \right) \right\|_F^2 \\
&= \mathcal{P}_{\text{box}} \left(G_{\hat{J}^c} \left(V R^{k+1} V^T - \frac{1}{\beta} (\hat{E} + Z^{k+\frac{1}{2}}) \right) \right),
\end{aligned} \tag{3.2}$$

where \mathcal{P}_{box} is the projection onto the polyhedral set $\{Y \in \mathbb{S}^{n_0+1} : 0 \leq Y \leq 1\}$.

3.2 Bounding

In this section we present some strategies for computing lower and upper bounds to **(IQP)**.

3.2.1 Lower Bounds from Lagrange Relaxation

We now discuss a strategy for computing a valid lower bound to $p_{\mathbf{IQP}}^*$. Exact solutions of the **DNN** relaxation (2.9) provide lower bounds to **(IQP)**. However, we often terminate algorithms when the stopping criteria are met for a pre-defined tolerance and we never set the tolerance to be exactly 0 in practice. A near optimal point \tilde{Y} can result in

$$p_{\mathbf{DNN}}^* \leq \langle \hat{E}, \tilde{Y} \rangle \text{ and } p_{\mathbf{IQP}}^* < \langle \hat{E}, \tilde{Y} \rangle$$

and produce an invalid lower bound to $p_{\mathbf{IQP}}^*$. Hence, we provide a method for computing a *valid lower bound* to **(IQP)** for avoiding this issue.

We follow the approaches in [12, 15, 23] and obtain lower bounds via the dual to the **DNN** relaxation in (2.9). We define the dual functional $g : \mathbb{S}^{n_0+1} \rightarrow \mathbb{R}$ by

$$g(Z) := \min_{R \in \mathcal{R}, Y \in \mathcal{Y}} \langle \hat{E}, Y \rangle + \langle Z, Y - V R V^T \rangle.$$

Let $\bar{Z} \in \mathbb{S}^{n_0+1}$ be given. We note that

$$\begin{aligned} \min_{R \in \mathcal{R}, Y \in \mathcal{Y}} \langle \hat{E}, Y \rangle + \langle \bar{Z}, Y - VRV^T \rangle &= \min_{Y \in \mathcal{Y}} \langle \hat{E} + \bar{Z}, Y \rangle + \min_{R \in \mathcal{R}} \langle -V^T \bar{Z} V, R \rangle \\ &= \min_{Y \in \mathcal{Y}} \langle \hat{E} + \bar{Z}, Y \rangle - (p+1) \lambda_{\max}(V^T \bar{Z} V), \end{aligned}$$

where λ_{\max} is the maximum eigenvalue function. Hence we compute a valid lower bound to the optimal value p_{DNN}^* of the model (2.9) by using weak duality:

$$p_{\text{DNN}}^* = \max_Z g(Z) \geq g(Z) = \min_{Y \in \mathcal{Y}} \langle \hat{E} + Z, Y \rangle - (p+1) \lambda_{\max}(V^T Z V),$$

where the first equality holds since the constraint qualification holds for the model (2.9). We note that the computation for $\min_{Y \in \mathcal{Y}} \langle \hat{E} + Z, Y \rangle$ is inexpensive.

3.2.2 Upper Bounds from Nearest Binary Feasible Solutions

In this section we discuss two strategies for computing upper bounds to the **SCP** problem. These strategies are derived from those presented in [5] and we include them here for completeness. We obtain upper bounds by finding feasible solutions to the original integer model in (2.3). Let $(R^{\text{out}}, Y^{\text{out}}, Z^{\text{out}})$ be the output of the algorithm.

1. Let $x^{\text{approx}} \in \mathbb{R}^{n_0}$ be the second through to the last elements of the first column of Y^{out} . Note that $0 \leq x^{\text{approx}} \leq 1$. Then the nearest feasible solution to (**IQP**) from x^{approx} can be found by solving the following projection:

$$\min_x \{ \|x - x^{\text{approx}}\|^2 : Ax = \bar{e}_p, x \in \{0, 1\}^{n_0} \}. \quad (3.3)$$

It is shown in [5] that solving (3.3) is equivalent to solving the following *linear program*:

$$\min_x \{ \langle x, x^{\text{approx}} \rangle : Ax = \bar{e}_p, x \geq 0 \}. \quad (3.4)$$

2. We now let x^{approx} be the second through to the last elements of the most dominant eigenvector of Y^{out} . Note that we again have $0 \leq x^{\text{approx}} \leq 1$, by the Perron-Frobenius theorem. We again obtain the nearest feasible solution to x^{approx} by solving (3.4).

Remark 3.2. *In fact, solving (3.4) does not require using any LP software; we can obtain the optimal solution for (3.4) as follows. We partition x^{approx} into p subvectors of sizes $m_i = |\mathcal{V}_i|$, for $i = 1, \dots, p$. Let $x^i \in \mathbb{R}^{m_i}$ be the subvector of x^{approx} associated with i -th rotamer set \mathcal{V}_i , i.e., $x^{\text{approx}} = [x^1; x^2; \dots; x^p]$. We define $\hat{x}^i \in \mathbb{R}^{m_i}$ as follows:*

$$\hat{x}_j^i = \begin{cases} 1, & \text{if } x_j^i = \max_{\ell \in [m_i]} \{x_\ell^i\} \\ 0, & \text{otherwise.} \end{cases}$$

*If there is subvector \hat{x}^i with more than one 1 in its components, we pick only one 1 and set the remaining to be 0. We then form $\hat{x} = [\hat{x}^1; \hat{x}^2; \dots; \hat{x}^p] \in \mathbb{R}^{n_0}$. It is clear that \hat{x} is feasible for (2.1). We use $\hat{x}^T E \hat{x}$ as an upper bound to the **SCP** problem.*

4 Numerical Experiments with Real-World Data

We present the numerical experiments for Algorithm 3.1. This section is organized as follows. In Section 4.1 we present the parameter settings and stopping criteria. In Section 4.2 we explain how we process the data from the Protein Data Bank (PDB) to obtain the energy matrix E . In Section 4.3 we finally present the numerical results using **rPRSM** and show that we provably solve many instances to optimality. We use the bounding strategies presented in Section 3.2 to prove optimality.

4.1 Stopping Criteria and Parameter Settings

Stopping Criteria We terminate **rPRSM** when either of the following conditions is satisfied.

1. Maximum number of iterations, denoted by “maxiter” is achieved.
2. For given tolerance ϵ , the following bound on the primal and dual residuals holds for s_t sequential times:

$$\max \left\{ \frac{\|Y^k - VR^kV^T\|_F}{\|Y^k\|_F} \beta \|Y^k - Y^{k-1}\|_F \right\} < \epsilon.$$

3. Let $\{l_1, \dots, l_k\}$ and $\{u_1, \dots, u_k\}$ be sequences of lower and upper bounds discussed in Section 3.2.1 and Section 3.2.2, respectively. Any of the lower bounds achieve the best upper bound, i.e.,

$$\min\{l_1, \dots, l_k\} \geq \max\{u_1, \dots, u_k\}.$$

Parameter Settings We use the following parameters related to the implementation of Algorithm 3.1:

$$\beta = \max\{[0.5 * n_0/p], 1\}, \quad \gamma = 0.99.$$

The parameters related to stopping criteria are:

$$\text{maxiter} = p(n_0 + 1) + 10^4, \quad \epsilon = 10^{-10}, \quad s_t = 100.$$

For the initial iterates for **rPRSM**, we use

$$Y^0 = 0, \quad Z^0 = \mathcal{P}_{Z_A}(Y^0).$$

4.2 Energy Matrix Computation

In this section we briefly describe the process for acquiring the energy matrix E . Our implementation relies on the usage of a Python script executing as an extension of the UCSF Chimera⁴ application. A detailed implementation can be found in [6, Chapter 7]. We used protein data files from the PDB to obtain the coordinates of all atoms in the protein. To get the energy values required by the algorithm, the native side chain conformations were replaced by rotamers extracted from a rotamer library provided by the Dunbrack Laboratory [11].

Some approaches use an energy evaluation based on a piece-wise linear approximation of the Lennard-Jones potential formula (e.g., [7, 28]). Here, we used the Lennard-Jones potential formula, which provides a more accurate energy value computation. In brief, the Lennard-Jones potential formula engages the Euclidean distance between a pair of atoms with some parameters dependant on the type of amino acids. A more detailed explanation of these energy computations can be found in [6, Chapter 6-7]. We finally used a strategy (known as ‘dead end elimination’) to reduce the size of the rotamer sets associated with each amino acid. The basic idea behind this strategy is that a rotamer can be removed from its rotamer set if there is another rotamer in that set that gives a better energy value regardless of the rotamer selections for the neighbouring amino acids. Among various approaches for the dead end elimination, we followed the Goldstein’s criteria [14].

Let \mathcal{U} be a side-chain conformation of a protein. The energy of the conformation \mathcal{U} is

$$E(\mathcal{U}) = \sum_{i=1}^{n_0} E_{\text{self}}(u_i) + \sum_{i=1}^{n_0-1} \sum_{j=i+1}^{n_0} E_{\text{pair}}(u_i, u_j),$$

⁴The UCSF Chimera software can be found in <https://www.cgl.ucsf.edu/chimera/download.html>.

where u_i is a side-chain conformation of an amino acid, $E_{\text{self}}(u_i)$ is the energy corresponding to u_i and the backbone, and $E_{\text{pair}}(u_i, u_j)$ is the energy formed by u_i and u_j , a rotamer associated with a neighbouring amino acid. In our formulation, we placed $E_{\text{self}}(u_i)$ along the diagonal of E and $E_{\text{pair}}(u_i, u_j)$ on the appropriate off-diagonal positions of E as shown in Section 2.1.

4.2.1 Removing Collisions

We typically observe some very large elements in E . This is due to the collisions between rotamers and they are indicated by huge values $E_{i,j} \gg 0$ that are often greater than 10^{10} . These huge values occur due to a part of the Lennard-Jones potential formula that involves the Euclidean distance between two distinct rotamers that goes to the denominator of a fraction.

In general, having very large values in data is prone to numerical instabilities. If every nonzero elements of E are large, the usual approach is to scale E to avoid large values. However, the matrix E often has elements that are more than 10 digits as well as elements that are 1 digit. When there is a large discrepancy among the elements of E , scaling E would make the relatively small values close to 0 and lead to loss of precision in the solution. However, this ill-posed data does not take place as a problem in our implementation. Recall that we update the Y iterate (3.2) as follows:

$$\begin{aligned} Y^{k+1} &= \mathcal{P}_Y \left(G_{\hat{J}^c} \left(VR^{k+1}V^T - \frac{1}{\beta}(\hat{E} + Z^{k+\frac{1}{2}}) \right) \right) \\ &= \mathcal{P}_Y \left(G_{\hat{J}^c} \left(-\frac{1}{\beta}\hat{E} + \left[VR^{k+1}V^T - \frac{1}{\beta}Z^{k+\frac{1}{2}} \right] \right) \right). \end{aligned}$$

For simplicity, we let $T := -\frac{1}{\beta}\hat{E} + \left[VR^{k+1}V^T - \frac{1}{\beta}Z^{k+\frac{1}{2}} \right]$. If the (\hat{i}, \hat{j}) -th element of $\hat{E} = \text{BlkDiag}(0, E)$ is very large, the projection \mathcal{P}_Y sets the (\hat{i}, \hat{j}) -element of T to 0 since $T_{\hat{i}, \hat{j}} \ll 0$. Hence, for those positions (\hat{i}, \hat{j}) with very large energy values, the constraint $Y_{\hat{i}, \hat{j}} = 0$ is implicitly imposed. We can interpret this as having implicit gangster constraints on these elements. Consequently, the large elements do not contribute to the objective value since $\hat{E}_{\hat{i}, \hat{j}} Y_{\hat{i}, \hat{j}} = 0$.

We can also take advantage of large values in the data to increase the number of the gangster indices (eliminate edges in the graph).

Lemma 4.1. *Suppose that x is feasible for (IQP), and let $u = x^T E x$ be its objective value. Let $N_E = \sum_{\{(i,j): E_{i,j} < 0\}} E_{i,j}$ and suppose that*

$$E_{i_0, j_0} > u - N_E, \text{ for some } i_0, j_0$$

holds. Then for any optimal solution x^ to (IQP), we have $x_{i_0}^* x_{j_0}^* = 0$.*

Proof. Let x^* be an optimal solution to (IQP). Let U^* be the set of indices formed by the positive entries of $\begin{pmatrix} 1 \\ x^* \end{pmatrix} \begin{pmatrix} 1 \\ x^* \end{pmatrix}^T$. We note that, for any index set S , we have

$$\sum_{(i,j) \in S} E_{i,j} = \sum_{(i,j) \in S \cap \{(i,j): E_{i,j} \geq 0\}} E_{i,j} + \sum_{(i,j) \in S \cap \{(i,j): E_{i,j} < 0\}} E_{i,j} \geq 0 + N_E = N_E.$$

Suppose to the contrary that x^* holds $x_{i_0}^* x_{j_0}^* = 1$, i.e., $x_{i_0}^* = x_{j_0}^* = 1$. Then we reach the following contradiction:

$$p_{\text{IQP}}^* = \langle x^*, E x^* \rangle = E_{i_0, j_0} + \left(E_{i_0, j_0} + \sum_{(i,j) \in U^* \setminus \{(i_0, j_0)\}} E_{i,j} \right) \geq E_{i_0, j_0} + N_E > u.$$

□

Corollary 4.2. Let i_0 be an index such that $E_{i_0, i_0} > u - N_E$, where u, N_E defined in Lemma 4.1. Then, for any optimal solution x^* to **(IQP)**, we have

$$Y_{x^*} := \begin{pmatrix} 1 \\ x^* \end{pmatrix} \begin{pmatrix} 1 \\ x^* \end{pmatrix}^T \in \{Y \in \mathbb{S}^{n_0+1} : Y(:, i_0) = 0, Y(i_0, :) = 0\}.$$

Proof. Let i_0 be an index such that $E_{i_0, i_0} > u - N_E$. Then $x_{i_0}^* = 0$ by Lemma 4.1. We note that Y_{x^*} is a positive semidefinite matrix. If a diagonal entry of a positive semidefinite is zero, then its corresponding column and row must be 0. \square

By Lemma 4.1 and Corollary 4.2, if we detect entries i_0, j_0 with the property $E_{i_0, j_0} > u - N_E$, then we may strengthen the model by adding the constraints

$$\mathcal{K} = \left\{ Y \in \mathbb{S}^{n_0+1} : \begin{array}{ll} Y(i_0, j_0) = Y(j_0, i_0) = 0, & \text{for } i_0 \neq j_0 \text{ such that } E_{i_0, j_0} > u - N_E \\ Y(:, i_0) = 0, Y(i_0, :) = 0, & \text{for } i_0 \text{ such that } E_{i_0, i_0} > u - N_E \end{array} \right\}.$$

This can be easily realized by adding more members to the gangster index set $\hat{\mathcal{J}}$.

4.3 Experiments with Real-World Data

In this section we provide numerical experiments with real-world data from Protein Data Bank and discuss the strengths of the **DNN** relaxation. We observe the useful aspects of the **DNN** relaxation through the numerical experiments. The **DNN** relaxation provides an effective treatment for avoiding numerical instabilities that originate from the large positive values in the data matrix E . Moreover, we observe that the **DNN** relaxation provides superior performance over the **SDP** relaxation.

We select instances listed in [7] with proteins that have up to 300 amino acids. All instances in Table 4.1 are tested using MATLAB version 2021a on Dell XPS 8940 with 11th Gen Intel(R) Core(TM) i5-11400 @ 2.60GHz 2.60 GHz with 32 Gigabyte memory. The following list defines the column headers used in Table 4.1; we use the same headers to the additional numerical experiments that are displayed in Appendix A.

1. **name**: instance name;
2. **p**: the number of amino acids;
3. **n₀**: the total number of rotamers;
4. **lbd**: the lower bound obtained by running **rPRSM**;
5. **ubd**: the upper bound obtained by running **rPRSM**;
6. **rel-gap**: relative gap of each instance using **rPRSM**, where

$$\text{relative gap} := 2 \frac{|\text{best feasible upper bound} - \text{best lower bound}|}{|\text{best feasible upper bound} + \text{best lower bound} + 1|};$$

7. **iter**: number of iterations used by **rPRSM** with tolerance $\epsilon = 10^{-10}$;
8. **time(sec)**: CPU time (in seconds) used by **rPRSM**.

| Problem Data | | | | Numerical Results | | | Timing | |
|--------------|------|-----|-------|-------------------|------------|-------------|--------|-----------|
| # | name | p | n_0 | lbd | ubd | rel-gap | iter | time(sec) |
| 10 | 2IGD | 50 | 126 | -78.50608 | -78.50608 | 5.39611e-15 | 500 | 19.43 |
| 20 | 1VQB | 75 | 406 | -96.94940 | -96.94940 | 4.34568e-14 | 900 | 179.35 |
| 30 | 2ACY | 84 | 580 | -146.32254 | -146.32254 | 1.06468e-14 | 7800 | 2610.24 |
| 40 | 2TGI | 100 | 355 | -14.03554 | -14.03554 | 2.46249e-13 | 1300 | 136.30 |
| 50 | 2SAK | 111 | 214 | -239.86975 | -239.86975 | 1.08995e-12 | 500 | 25.50 |
| 60 | 2CPL | 132 | 819 | -284.97180 | -284.97180 | 9.75693e-15 | 5900 | 3292.98 |
| 70 | 1CV8 | 146 | 730 | -213.13554 | -213.13554 | 3.28738e-13 | 5600 | 2572.99 |
| 80 | 2ENG | 162 | 867 | 82.01797 | 82.01797 | 1.33295e-13 | 14200 | 8274.48 |
| 90 | 1A7S | 179 | 524 | -239.78218 | -239.78218 | 1.00542e-14 | 1200 | 314.57 |
| 100 | 1MRJ | 208 | 1178 | -295.13711 | -295.13711 | 1.70740e-13 | 2300 | 2421.15 |
| 110 | 1EZM | 239 | 1497 | -217.36581 | -217.36581 | 3.49620e-13 | 2300 | 3876.18 |
| 120 | 1SBP | 256 | 1704 | -271.08838 | -271.08838 | 3.59996e-14 | 40000 | 609487.29 |
| 130 | 3PTE | 284 | 2006 | 161.17216 | 161.17216 | 5.09815e-15 | 13500 | 250604.17 |

Table 4.1: Computational results on selected PDB instances

Discussion We observe from the last two columns of Table 4.1 that many instances are solved within good relative gaps. In fact, most of the instances display relative gaps that are essentially 0. We recall from (3.4) that we obtain the upper bounds via finding feasible solutions to (IQP). We recall from Section 3.2.2 that we obtain the upper bounds via finding feasible solutions to (IQP). That we have the relative gap essentially 0 grants us the attainment of the *globally optimal* solutions to the SCP problem. Approaches involving heuristic algorithms do not provide a natural means of certifying optimality, relying solely on a comparison of the rotameric solution with naive χ_1 and χ_2 angles from the PDB while ignoring optimality of the discretized solution. We highlight that we provide not only the globally optimal solutions but also a way to certify their optimality.

4.3.1 A Tighter Relaxation

We illustrate the strengths of the DNN relaxation by computing the near optimal values of the DNN relaxation and the SDP relaxation. In our test, we selected five small instances. As discussed above, some elements of the energy matrix E are typically very large due to the collisions in rotamers, typically at least 10 digits. These cause numerical difficulties when a standard interior point solver is used. Hence, in our test, we set the entries $E_{i,j} = \min\{10^4, E_{i,j}\}$, $\forall i, j$, in order to avoid the difficulties from having these large elements. We used the rPRSM for DNN relaxation and used SDPT3⁵ for solving the SDP relaxation. The displayed values in

| problem # | instance | DNN relaxation | SDP relaxation |
|-----------|----------|----------------|----------------|
| 1 | 1AIE | -46.96 | -2460.53 |
| 2 | 2ERL | 55.33 | -18241.26 |
| 3 | 1CBN | -40.43 | -22380.58 |
| 4 | 1RB9 | -76.97 | -23936.35 |
| 5 | 1BX7 | 16.96 | -23965.88 |

Table 4.2: The solver optimal values of the DNN and SDP relaxations on selected instances

Table 4.2 are the best lower bounds found from the rPRSM and the optimal values reported by SDPT3. We observe in Table 4.2 that the DNN relaxation shows superior performances over the

⁵<https://www.math.cmu.edu/~reha/sdpt3.html>, version SDPT3 4.0, [26].

SDP relaxations in the relaxation values; the **DNN** relaxation for the **SDP** problem provides a much tighter relaxation than the **SDP** relaxation.

5 Conclusions

We presented a simple way of formulating the relaxation of the **SCP** problem. We began by formulating the **SCP** problem into an **IQP** and derived the facially reduced **SDP** relaxation. We then identified some redundant constraints to the **IQP** to complete the **DNN** relaxation. **FR** allowed for a natural splitting of the variables and provided a perfect environment for using splitting methods. Hence we adopted the **rPRSM** to solve the **DNN** relaxation of the **SCP** problem. We illustrated the efficiency of our approach using data from the Protein Data Bank. In particular, we solved many instances chosen from the Protein Data Bank to optimality.

Index

- Y_{ij}^{col} , 8
- A , 5
- E , edge weights, 4
- E_{00} , 5
- $G_{\mathcal{J}}$, gangster operator, 6
- $G_{\hat{\mathcal{J}}}$, gangster operator, 6
- K , 7
- N_E , 14
- $X \succ 0$, 3
- $X \succeq 0$, 3
- Y_x , 5
- $[m] = \{1, \dots, m\}$, 3
- $\langle \cdot, \cdot \rangle$, 3
- BlkDiag, 5
- BlkDiag, 3
- Diag, 3
- $\mathbb{R}^{m \times n}$, 3
- \mathbb{R}^n , 3
- \mathbb{S}_+^n , 3
- \mathbb{S}_{++}^n , 3
- \bar{e}_n , 3
- \mathcal{J} , 6
- $\mathcal{L}_A(R, Y, Z)$, augmented Lagrangian, 9
- $\mathcal{N}_{\mathcal{C}}$, 3
- $\mathcal{P}_{Z_0}(Z)$, 10
- \mathcal{P}_{box} , 11
- \mathcal{R} , 9
- \mathcal{V} , 4
- \mathcal{V}_i , i -th rotamer set, 4
- \mathcal{Y} , 9
- diag, 3
- \hat{E} , 5
- $\hat{\mathcal{J}}$, 6
- λ_{\max} , 11
- e_0 , the first unit vector, 5
- g , dual functional, 11
- i -th rotamer set, \mathcal{V}_i , 4
- m_i , 4
- n_0 , total number of rotamers, 4, 5
- p_{DNN}^* , optimal value of **DNN** relaxation, 9
- p_{IQP}^* , optimal value of **IQP**, 4
- s_t , 12
- $\mathcal{V} = \cup_{i=1}^p \mathcal{V}_i$, 4
- \mathcal{Z}_A , 9
- DNN**, doubly nonnegative relaxation, 2, 3
- FR**, facial reduction, 3
- IQP**, integer quadratic problem, 2
- PRSM**, Peaceman-Rachford splitting method, 10
- SCP**, side-chain positioning, 2, 4
- SDP**, semidefinite programming, 3
- DNN**, 9
- IQP**, 5
- SDP**, 5
- augmented Lagrangian, $\mathcal{L}_A(R, Y, Z)$, 9
- collisions, 14
- dead end elimination, 13
- doubly nonnegative (**DNN**), 3
- dual functional, g , 11
- edge weights, E , 4
- exposing vector, 7
- facial reduction, 7
- facial reduction (**FR**), 3
- gangster constraint, 6
- gangster operator, $G_{\mathcal{J}}$, 6
- gangster operator, $G_{\hat{\mathcal{J}}}$, 6
- integer quadratic problem (**IQP**), 3
- normal cone, 3
- optimal value of **DNN** relaxation, p_{DNN}^* , 9
- optimal value of **IQP**, p_{IQP}^* , 4
- PDB, Protein Data Bank, 12
- Peaceman-Rachford splitting method (**PRSM**), 2, 3
- Protein Data Bank, PDB, 12
- protein side-chain positioning problem, 4
- restricted dual Peaceman-Rachford splitting method, **rPRSM**, 10
- rotamer, 4
- semidefinite programming (**SDP**), 3
- side-chain positioning (**SCP**), 2
- the first unit vector, e_0 , 5
- total number of rotamers, n_0 , 5
- trace, 3

References

- [1] T. Akutsu. Np-hardness results for protein side-chain packing. *Genome Informatics*, 8:180–186, 1997. [3](#)
- [2] E. Althaus, O. Kohlbacher, H.-P. Lenhof, and P. Mauller. A combinatorial approach to protein docking with flexible side chains. *Journal of computational biology*, 9(4):597–612, 2002. [3](#)
- [3] D. Bahadur, T. Akutsu, E. Tomita, and T. Seki. Protein side-chain packing problem: A maximum edge-weight clique algorithmic approach. *Journal of bioinformatics and computational biology*, 3 1:103–26, 2004. [3](#)
- [4] M.J. Bower, F.E. Cohen, and R.L. Dunbrack. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *Journal of Molecular Biology*, 267(5):1268–1282, 1997. [3](#)
- [5] F. Burkowski, Y-L. Cheung, and H. Wolkowicz. Efficient use of semidefinite programming for selection of rotamers in protein conformations. *INFORMS J. Comput.*, 26(4):748–766, 2014. [3](#), [5](#), [7](#), [9](#), [10](#), [12](#)
- [6] F.J. Burkowski. *Computational and Visualization Techniques for Structural Bioinformatics Using Chimera*. Chapman & Hall/CRC mathematical and computational biology series. Chapman and Hall/CRC, London, 2015. [13](#)
- [7] A.A. Canutescu, A.A. Shelenkov, and R.L. Dunbrack. A graph-theory algorithm for rapid protein side-chain prediction. *Protein science*, 12(9):2001–2014, 2003. [3](#), [13](#), [15](#)
- [8] B. Chazelle, C. Kingsford, and M. Singh. A semidefinite programming approach to side chain positioning with new rounding strategies. *INFORMS J. Comput.*, 16(4):380–392, 2004. [3](#)
- [9] J. Desmet, M. De Maeyer, B. Hazes, and I. Lasters. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature (London)*, 356(6369):539–542, 1992. [3](#)
- [10] D. Drusvyatskiy and H. Wolkowicz. The many faces of degeneracy in conic optimization. *Foundations and Trends[®] in Optimization*, 3(2):77–170, 2017. [7](#)
- [11] R.L. Dunbrack, Jr. and M. Karplus. Backbone-dependent rotamer library for proteins application to side-chain prediction. *Journal of Molecular Biology*, 230(2):543–574, March 1993. [13](#)
- [12] J. Eckstein. Deriving solution value bounds from the ADMM. *Optimization Letters*, 2020. [11](#)
- [13] O. Eriksson, Y. Zhou, and A. Elofsson. Side chain-positioning as an integer programming problem. In *Algorithms in bioinformatics (Århus, 2001)*, volume 2149 of *Lecture Notes in Comput. Sci.*, pages 128–141. Springer, Berlin, 2001. [3](#)
- [14] R.F. Goldstein. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophysical Journal*, 66(5):1335 – 1340, 1994. [13](#)
- [15] N. Graham, H. Hu, H. Im, X. Li, and H. Wolkowicz. A restricted dual Peaceman-Rachford splitting method for a strengthened DNN relaxation for QAP. *INFORMS J. Comput.*, 34(4):2125–2143, 2022. [3](#), [9](#), [10](#), [11](#)

- [16] L. Holm and C. Sander. Database algorithm for generating protein backbone and side-chain co-ordinates from a c it trace : Application to model building and detection of co-ordinate errors. *Journal of Molecular Biology*, 218(1):183–194, 1991. [3](#)
- [17] C.L Kingsford, B. Chazelle, and M. Singh. Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics (Oxford, England)*, 21(7):1028–1039, 2005. [3](#)
- [18] V. Laudet and H. Gronemeyer. 3 - ligand binding. In V. Laudet and H. Gronemeyer, editors, *The Nuclear Receptor FactsBook*, Factsbook, pages 37 – 41. Academic Press, London, 2002. [2](#)
- [19] C. Lee. Predicting protein mutant energetics by self-consistent ensemble optimization. *Journal of Molecular Biology*, 236(3):918–939, 1994. [3](#)
- [20] X. Li, T.K. Pong, H. Sun, and H. Wolkowicz. A strictly contractive Peaceman-Rachford splitting method for the doubly nonnegative relaxation of the minimum cut problem. *Comput. Optim. Appl.*, 78(3):853–891, 2021. [3](#)
- [21] L.L Looger, M.A Dwyer, J.J Smith, and H.W Hellinga. Computational design of receptor and sensor proteins with novel functions. *Nature (London)*, 423(6936):185–190, 2003. [2](#)
- [22] N.A. Marze, S.S. Roy-Burman, W. Sheffler, and J.J. Gray. Efficient flexible backbone protein-protein docking for challenging targets. *Computer applications in the biosciences*, 34(20):3461–3469, 2018. [2](#)
- [23] D.E. Oliveira, H. Wolkowicz, and Y. Xu. ADMM for the SDP relaxation of the QAP. *Math. Program. Comput.*, 10(4):631–658, 2018. [3](#), [11](#)
- [24] R. Samudrala and J. Moult. Determinants of side chain conformational preferences in protein structures. *Protein engineering*, 11(11):991–997, 1998. [3](#)
- [25] P.S. Shenkin, H. Farid, and J.S. Fetrow. Prediction and evaluation of side-chain conformations for protein backbone structures. *Proteins: Structure, Function, and Bioinformatics*, 26(3):323–352, 1996. [3](#)
- [26] K.C. Toh, M.J. Todd, and R.H. Tütüncü. SDPT3—a MATLAB software package for semidefinite programming, version 1.3. *Optim. Methods Softw.*, 11/12(1-4):545–581, 1999. Interior point methods. [16](#)
- [27] C. Wang, P. Bradley, and D. Baker. Protein-protein docking with backbone flexibility. *Journal of molecular biology*, 373(2):503–519, 2007. [2](#)
- [28] J. Xu and B. Berger. Fast and accurate algorithms for protein side-chain packing. *Journal of the ACM (JACM)*, 53(4):533–557, 2006. [3](#), [13](#)

A Additional Numerics

| Problem Data | | | | Numerical Results | | | Timing | |
|--------------|------|-----|-------|-------------------|------------|-------------|--------|-----------|
| # | name | p | n_0 | lbd | ubd | rel-gap | iter | time(sec) |
| 1 | 1AIE | 26 | 34 | -46.95892 | -46.95892 | 1.04802e-15 | 200 | 0.10 |
| 2 | 2ERL | 34 | 103 | 55.33285 | 55.33284 | 1.17985e-12 | 200 | 5.85 |
| 3 | 1CBN | 37 | 112 | -40.42751 | -40.42751 | 1.68402e-14 | 300 | 7.77 |
| 4 | 1RB9 | 41 | 105 | -76.96501 | -76.96501 | 7.11964e-13 | 1000 | 26.39 |
| 5 | 1BX7 | 41 | 99 | 16.96026 | 16.96026 | 5.21525e-12 | 300 | 7.25 |
| 6 | 2FDN | 42 | 51 | -59.43091 | -59.43092 | 3.71094e-14 | 200 | 0.04 |
| 7 | 1MOF | 46 | 94 | -79.05580 | -79.05580 | 3.52629e-12 | 200 | 4.03 |
| 8 | 1CTF | 47 | 74 | -97.18893 | -97.18893 | 4.64633e-13 | 200 | 2.81 |
| 9 | 1NKD | 50 | 199 | -51.78466 | -51.78466 | 4.40639e-12 | 2680 | 192.65 |
| 10 | 2IGD | 50 | 126 | -78.50608 | -78.50608 | 5.39611e-15 | 500 | 14.67 |
| 11 | 2SN3 | 53 | 112 | -5.56818 | -5.56818 | 6.73872e-13 | 700 | 16.77 |
| 12 | 1MSI | 54 | 112 | -87.46958 | -87.46958 | 1.72043e-13 | 700 | 19.39 |
| 13 | 1AHO | 54 | 140 | 24.66925 | 24.66925 | 4.19224e-14 | 1500 | 56.22 |
| 14 | 1COR | 60 | 131 | 15.58314 | 15.58314 | 4.58637e-12 | 1000 | 32.31 |
| 15 | 1CTJ | 61 | 258 | -103.32705 | -103.32705 | 1.64217e-12 | 1872 | 162.80 |
| 16 | 1RZL | 65 | 121 | 17.26470 | 17.26470 | 1.22992e-11 | 2468 | 68.52 |
| 17 | 1TIF | 66 | 614 | -155.17859 | -155.17859 | 4.69196e-14 | 1000 | 350.89 |
| 18 | 1BDO | 69 | 221 | -136.29933 | -136.29933 | 8.93377e-15 | 1000 | 75.06 |
| 19 | 1OPD | 70 | 112 | -139.64632 | -139.64632 | 1.18233e-13 | 300 | 5.98 |
| 20 | 1VQB | 75 | 406 | -96.94940 | -96.94940 | 4.34568e-14 | 900 | 147.36 |
| 21 | 1IUZ | 75 | 221 | -150.88238 | -150.88238 | 1.25791e-14 | 3200 | 227.45 |
| 22 | 1ABA | 76 | 376 | -137.59962 | -137.59963 | 9.05546e-15 | 600 | 88.43 |
| 23 | 1FNA | 76 | 131 | -172.01313 | -172.01313 | 3.64100e-14 | 800 | 23.32 |
| 24 | 1CYO | 78 | 220 | -75.36668 | -75.36668 | 1.36739e-14 | 700 | 48.50 |
| 25 | 1FUS | 79 | 302 | -4.66627 | -4.66627 | 1.11145e-12 | 3000 | 312.35 |
| 26 | 2MCM | 80 | 123 | -135.14024 | -135.14024 | 8.30816e-13 | 400 | 10.30 |
| 27 | 1SVY | 80 | 147 | -141.92437 | -141.92437 | 6.21219e-13 | 400 | 14.51 |
| 28 | 1A68 | 81 | 424 | -178.12555 | -178.12555 | 2.54581e-15 | 1500 | 249.80 |
| 29 | 1YCC | 84 | 223 | -79.21270 | -79.21270 | 2.11079e-12 | 955 | 66.26 |
| 30 | 2ACY | 84 | 580 | -146.32254 | -146.32254 | 1.06468e-14 | 7800 | 2175.04 |
| 31 | 1BM8 | 85 | 687 | -119.54537 | -119.54537 | 2.02428e-14 | 1300 | 509.88 |
| 32 | 1BKF | 89 | 339 | -170.80514 | -170.80514 | 1.60935e-14 | 1000 | 117.73 |
| 33 | 3CYR | 91 | 137 | -144.06405 | -144.06405 | 2.48290e-12 | 1900 | 52.09 |
| 34 | 3VUB | 92 | 544 | -229.38312 | -229.38312 | 7.41813e-16 | 1400 | 349.67 |
| 35 | 1JER | 96 | 462 | -120.78401 | -120.78400 | 1.15131e-12 | 3232 | 633.90 |
| 36 | 2HBG | 97 | 275 | -178.42210 | -178.42210 | 2.70839e-13 | 500 | 42.98 |
| 37 | 1POA | 97 | 470 | 278.08280 | 278.08280 | 2.02964e-12 | 5463 | 1099.55 |
| 38 | 1C52 | 99 | 256 | -223.31096 | -223.31096 | 2.41281e-15 | 2700 | 203.46 |
| 39 | 2A0B | 99 | 642 | -161.45228 | -161.45228 | 1.75494e-16 | 5200 | 1800.90 |
| 40 | 2TGI | 100 | 355 | -14.03554 | -14.03554 | 2.46249e-13 | 1300 | 153.95 |

Table A.1: Computation results on selected PDB instances up to 100 amino acids

| Problem Data | | | | Numerical Results | | | Timing | |
|--------------|------|-----|-------|-------------------|---------------|-------------|--------|-----------|
| # | name | p | n_0 | lbd | ubd | rel-gap | iter | time(sec) |
| 41 | 3NUL | 101 | 285 | -154.87542 | -154.87542 | 1.28046e-15 | 2300 | 307.34 |
| 42 | 1WHI | 101 | 298 | -247.13457 | -247.13457 | 6.94375e-14 | 1500 | 199.52 |
| 43 | 1PDO | 104 | 453 | -188.29848 | -188.29848 | 9.10541e-12 | 5754 | 1456.33 |
| 44 | 3LZT | 105 | 530 | -48.81821 | -48.81821 | 8.48591e-13 | 1100 | 300.50 |
| 45 | 1DHN | 105 | 519 | -133.77464 | -133.77464 | 1.35468e-13 | 2000 | 535.83 |
| 46 | 1KUH | 106 | 580 | -155.56590 | -155.56590 | 2.18536e-15 | 2296 | 743.57 |
| 47 | 1ECA | 108 | 655 | -169.74717 | -169.74717 | 1.66944e-16 | 25200 | 12563.89 |
| 48 | 1BFG | 108 | 410 | -191.73261 | -191.73262 | 8.54577e-14 | 900 | 210.84 |
| 49 | 1RIE | 108 | 930 | -117.91809 | -117.91809 | 1.57208e-14 | 20200 | 17809.01 |
| 50 | 2SAK | 111 | 214 | -239.86975 | -239.86975 | 1.08995e-12 | 500 | 37.26 |
| 51 | 1BGF | 112 | 1180 | -239.65571 | -239.65571 | 1.52549e-13 | 56400 | 71503.54 |
| 52 | 2END | 118 | 707 | -8.22833 | -8.22833 | 1.08596e-12 | 16100 | 8511.24 |
| 53 | 2SNS | 119 | 634 | 620.86546 | 620.86546 | 1.79304e-14 | 6900 | 3082.12 |
| 54 | 1BD8 | 121 | 347 | -219.12419 | -219.12419 | 9.42666e-12 | 4970 | 760.81 |
| 55 | 1NPK | 122 | 709 | -205.56059 | -205.56059 | 6.77231e-13 | 59075 | 31212.37 |
| 56 | 1A6M | 124 | 613 | -55.41007 | -55.41008 | 4.93096e-14 | 22800 | 7608.82 |
| 57 | 2RN2 | 127 | 830 | -198.37189 | -198.37189 | 1.41057e-13 | 6073 | 4053.13 |
| 58 | 1RCF | 130 | 733 | -86.59895 | -86.59775 | 1.38011e-05 | 100000 | 56927.20 |
| 59 | 1LCL | 131 | 1246 | -217.16433 | -217.16433 | 2.53317e-14 | 3800 | 4821.11 |
| 60 | 2CPL | 132 | 819 | -284.97180 | -284.97180 | 9.75693e-15 | 5900 | 3329.39 |
| 61 | 1VHH | 133 | 844 | -21.33604 | -21.33604 | 3.59566e-14 | 3200 | 1843.96 |
| 62 | 1BJ7 | 135 | 917 | -64.37915 | -64.37915 | 5.69493e-14 | 11300 | 8946.94 |
| 63 | 119L | 136 | 970 | -234.21535 | -234.21535 | 8.01617e-14 | 34200 | 30890.87 |
| 64 | 1RA9 | 136 | 1018 | -185.07235 | -185.07235 | 5.13076e-14 | 4400 | 4839.16 |
| 65 | 1L58 | 137 | 962 | -285.60167 | -285.60167 | 1.31131e-14 | 15600 | 13812.60 |
| 66 | 2ILK | 142 | 708 | -121.02712 | -121.02712 | 1.82770e-13 | 4700 | 2750.13 |
| 67 | 1KOE | 144 | 710 | -13.87537 | -13.87537 | 1.27269e-11 | 4124 | 2490.08 |
| 68 | 1HA1 | 146 | 538 | -213.93793 | -213.93793 | 1.44469e-13 | 3700 | 1229.31 |
| 69 | 1CEX | 146 | 415 | 174.95279 | 174.95279 | 2.40438e-11 | 11447 | 2426.49 |
| 70 | 1CV8 | 146 | 730 | -213.13554 | -213.13554 | 3.28738e-13 | 5600 | 3442.13 |
| 71 | 153L | 149 | 846 | -170.13061 | -170.13061 | 3.03488e-13 | 2100 | 1554.46 |
| 72 | 1BS9 | 150 | 935 | 103.16569 | 103.16569 | 1.31052e-13 | 2500 | 1736.57 |
| 73 | 2PTH | 151 | 1198 | -190.97344 | -190.97344 | 1.39085e-13 | 1900 | 2233.17 |
| 74 | 1XNB | 151 | 1233 | -147.30040 | -147.30040 | 2.69217e-15 | 13300 | 16562.76 |
| 75 | 1AQB | 152 | 713 | 29.24537 | 29.24537 | 9.30418e-14 | 39300 | 17795.39 |
| 76 | 1LBU | 152 | 1225 | 38.14603 | 38.14603 | 1.91397e-13 | 9900 | 11673.18 |
| 77 | 1KID | 153 | 653 | -351.91160 | -351.91160 | 2.90337e-15 | 6600 | 2607.24 |
| 78 | 1CHD | 154 | 489 | -164.21510 | -164.21510 | 3.27846e-14 | 19300 | 4097.50 |
| 79 | 1AMM | 158 | 1480 | -288.62671 | -288.62671 | 2.75245e-15 | 3300 | 5793.13 |
| 80 | 2ENG | 162 | 867 | 82.01797 | 82.01797 | 1.33295e-13 | 14200 | 8284.65 |
| 81 | 1G3P | 165 | 921 | -70.30769 | -70.30769 | 6.66312e-14 | 7000 | 4469.99 |
| 82 | 1THV | 167 | 902 | 5.12749 | 5.12749 | 4.63732e-12 | 4200 | 2637.88 |
| 83 | 1PPN | 170 | 1259 | -56.69346 | -56.69346 | 1.23365e-13 | 11589 | 14139.22 |
| 84 | 1IAB | 173 | 775 | 321.20652 | 321.20652 | 2.04964e-14 | 26500 | 13017.74 |
| 85 | 1DIN | 175 | 1110 | -264.73564 | -264.73548 | 5.84356e-07 | 100000 | 93357.26 |
| 86 | 2AYH | 176 | 1269 | 8428.18154 | 6089367.83709 | 1.99447e+00 | 100000 | 135879.29 |
| 87 | 1ZIN | 177 | 853 | -353.00431 | -353.00431 | 3.18384e-14 | 23800 | 13742.52 |
| 88 | 1BYI | 177 | 818 | -242.78881 | -242.78881 | 2.33646e-14 | 2400 | 1298.65 |
| 89 | 2BAA | 178 | 1165 | -43.77265 | -43.77265 | 1.95480e-12 | 4600 | 4785.88 |
| 90 | 1A7S | 179 | 524 | -239.78218 | -239.78218 | 1.00542e-14 | 1200 | 284.88 |
| 91 | 1WAB | 183 | 1063 | -317.46713 | -317.46713 | 9.40337e-14 | 8500 | 7357.75 |
| 92 | 1MUN | 185 | 1047 | -378.01261 | -378.01261 | 1.15635e-14 | 9500 | 7883.00 |
| 93 | 1LST | 192 | 946 | -244.76861 | -244.76861 | 1.28627e-14 | 32300 | 21374.44 |
| 94 | 1GCI | 194 | 1052 | -205.63185 | -205.63185 | 2.79899e-14 | 10300 | 8885.03 |
| 95 | 3CLA | 198 | 857 | -26.72768 | -26.72768 | 9.89051e-14 | 3900 | 2287.99 |

Table A.2: Computation results on selected PDB instances up to 200 amino acids

| Problem Data | | | | Numerical Results | | | Timing | |
|--------------|------|-----|-------|-------------------|------------|-------------|--------|-----------|
| # | name | p | n_0 | lbd | ubd | rel-gap | iter | time(sec) |
| 96 | 1AL3 | 201 | 1077 | 119.66598 | 119.66598 | 3.39407e-14 | 12500 | 10188.87 |
| 97 | 1ARB | 202 | 1466 | -61.52823 | -61.52823 | 3.41363e-14 | 8900 | 14632.82 |
| 98 | 1XJO | 202 | 776 | -171.92443 | -171.92443 | 8.24179e-15 | 3700 | 1455.50 |
| 99 | 1NLS | 203 | 1060 | -297.73578 | -297.73578 | 5.33677e-15 | 2500 | 1976.08 |
| 100 | 1MRJ | 208 | 1178 | -295.13711 | -295.13711 | 1.70740e-13 | 2300 | 2149.63 |
| 101 | 1OAA | 208 | 854 | -317.83422 | -317.83422 | 1.44174e-12 | 3842 | 1823.52 |
| 102 | 2DRI | 210 | 906 | -398.45564 | -398.45564 | 2.56465e-15 | 6200 | 3225.99 |
| 103 | 2CBA | 223 | 1018 | -86.52145 | -86.52145 | 5.34000e-14 | 3400 | 2407.24 |
| 104 | 2POR | 224 | 1304 | -83.22221 | -83.22221 | 5.55044e-14 | 6700 | 8044.39 |
| 105 | 3SEB | 224 | 1412 | 77.15838 | 77.15852 | 1.84867e-06 | 100000 | 137194.81 |
| 106 | 1MLA | 227 | 1322 | -484.10542 | -484.10542 | 1.68910e-14 | 62900 | 75257.79 |
| 107 | 1DCS | 232 | 1170 | -342.68600 | -342.68600 | 1.39133e-14 | 8000 | 7459.07 |
| 108 | 1AKO | 234 | 1387 | -244.65691 | -244.65691 | 1.18251e-14 | 7400 | 9809.00 |
| 109 | 1PDA | 239 | 891 | -423.50226 | -423.50226 | 4.96037e-15 | 9100 | 4520.68 |
| 110 | 1EZM | 239 | 1497 | -217.36581 | -217.36581 | 3.49620e-13 | 2300 | 3919.92 |
| 111 | 1C3D | 243 | 1679 | -400.69876 | -400.69876 | 1.04846e-14 | 22100 | 134094.53 |
| 112 | 1RHS | 244 | 1973 | -341.20443 | -341.20443 | 1.41400e-14 | 7300 | 62136.57 |
| 113 | 8ABP | 245 | 1743 | -273.90715 | -273.90716 | 2.27865e-15 | 9000 | 59868.98 |
| 114 | 1CVL | 246 | 910 | -537.04249 | -537.04249 | 2.11494e-16 | 14800 | 7522.51 |
| 115 | 1RYC | 248 | 1831 | -202.60568 | -202.60568 | 4.81378e-14 | 15200 | 84674.22 |
| 116 | 1MRP | 248 | 1648 | -350.97062 | -350.97062 | 1.39088e-14 | 11000 | 34303.23 |
| 117 | 1IXH | 252 | 1134 | -289.75241 | -289.75241 | 4.11267e-14 | 1300 | 1087.30 |
| 118 | 1FNC | 253 | 1940 | -310.60999 | -310.60999 | 6.54656e-13 | 34321 | 292924.91 |
| 119 | 1TCA | 255 | 1062 | -422.15387 | -422.15387 | 4.24994e-14 | 8700 | 6424.87 |
| 120 | 1SBP | 256 | 1704 | -271.08838 | -271.08838 | 3.59996e-14 | 40000 | 156330.60 |
| 121 | 2CTC | 264 | 1536 | -213.88596 | -213.88596 | 2.17419e-14 | 15100 | 43642.85 |
| 122 | 1PGS | 265 | 2190 | -16.14049 | -16.14049 | 2.28785e-12 | 21300 | 269611.15 |
| 123 | 1MSK | 271 | 1798 | -162.51007 | -162.50978 | 1.77573e-06 | 100000 | 771330.61 |
| 124 | 1BG6 | 271 | 784 | -452.62383 | -452.62383 | 3.13620e-15 | 12700 | 4935.11 |
| 125 | 1ARU | 271 | 939 | -314.40612 | -314.40589 | 7.15908e-07 | 100000 | 53858.54 |
| 126 | 1A8E | 274 | 1096 | -249.85499 | -249.85499 | 3.58741e-14 | 96500 | 78746.74 |
| 127 | 1AXN | 278 | 2343 | -300.34291 | -300.34291 | 7.55789e-15 | 12500 | 207625.02 |
| 128 | 1TAG | 279 | 1330 | -253.22167 | -253.22167 | 1.68029e-14 | 4300 | 5038.43 |
| 129 | 1ADS | 280 | 1560 | 733.91439 | 733.91440 | 1.39319e-13 | 18273 | 65301.22 |
| 130 | 3PTE | 284 | 2006 | 161.17216 | 161.17216 | 5.09815e-15 | 13500 | 59169.60 |
| 131 | 1CEM | 292 | 2400 | -24.20196 | -24.20196 | 3.85446e-14 | 7000 | 47701.70 |

Table A.3: Computation results on selected PDB instances up to 300 amino acids