Facial reduction for semidefinite programming and its application for the selection of rotamers in protein conformations

Henry Wolkowicz

Dept. Combinatorics and Optimization,
University of Waterloo

with: Forbes Burkowski and Yuen-Lam (Cheung) Voronin

Outline: Modeling/Degeneracy in SDP Relaxation

- Model the NP-hard side chain positioning problem using a QQP: quadratic (objective) - quadratic (constraints) program
- Find the standard semidefinite (SDP) relaxation for the QQP
- show: SDP relaxation is degenerate (not strictly feasible) (causes problems in theory and numerics)
- Preprocess/regularize using **facial reduction**
  - -two types of facial reduction
  - -facial reduction improves/strengthens numerics
- strengthen solutions using redundant quadratic constraints in model and using cutting plane techniques

We follow/improve/strengthen SDP relaxation approaches in:
-chazelle, Kingsford, Singh for SCP, 2004
-Qing, Karish, Rendl, W. for QAP, 1998.

Biological Preamble I

Side chain positioning (SCP)

- Given: constituent atoms of a protein;
  the side chain positioning (SCP) problem
  is one of the multiple subproblems of the hard problem of predicting a protein's three dimensional structure.
- Our protein macromolecule is a chain of amino acids, also called reisdues.

Amino acid is characterized by composition of its side chain

- amino acid consists of an "alpha" carbon atom (-C_α-), and three components attached to it:
  -(i) amino group ($(H_2N$-);
  -(ii) carboxyl group (-COOH);
  -(iii) atom group called a side chain

Biological Preamble II

Backbone of the protein

- Atoms in the *backbone* of the protein form a repetitive sequence of triplets: $\cdots \mathrm{NC_\alpha C} \quad \mathrm{NC_\alpha C} \quad \mathrm{NC_\alpha C} \quad \mathrm{NC_\alpha C} \cdots$ with each CN bonding being the result of a condensation reaction.
- Protein chain is a repetitive sequence of atoms with side chain groups sprouting from the alpha carbon atoms.

Famous protein folding problem

Outline:

For tractability, accurate prediction of all atomic positions for folded minimal energy conformation typically uses:

1. calculate the positions of atoms in the backbone (e.g., homology modeling; fold recognition techniques)
2. given the positions of backbone atoms, calculate the conformations of all side chains, SCP.

Rotamericity/discretization of side chain conformations

- side chain typically adopts a conformation close to one of finitely many possible dihedral angles; each of the finite number of three dimensional conformations is called a rotamer.
- In this work: our more complicated side chains have rotamer sets with as many as 81 members for the twenty amino acids that make up proteins.

Modelling

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, E)$ weighted, undirected graph

- node set $\mathcal{V} = \bigcup_{i=1}^{p} \mathcal{V}_i$, $\mathcal{V}_i$ subset of rotamers for i-th amino acid side chain/residue position, p is the number of residues.
- edge set $\mathcal{E}$; weights (energy between rotamers) E_{uv} for edge $uv \cong (u, v) \in \mathcal{E}$; E_{uu} is energy between backbone and chosen rotamer u. (ref. Kingsford thesis)

Further: SDP notation

- $\mathcal{S}^t$, $t \times t$ real symmetric matrices, trace inner-product $\langle S, T \rangle = \text{trace}\, ST$; Löwner partial order $S \succeq T$, $S \succ T$.
- for $v \in \mathbb{R}^s$, corresp. diagonal matrix is $\text{Diag}\,(v) \in \mathcal{S}^s$ adjoint linear transformation is $\text{Diag}^*(S) = \text{diag}\,(S) \in \mathbb{R}^s$ the adjoint satisfies $\langle \text{diag}\,(S), v \rangle = \langle S, \text{Diag}\,(v) \rangle$
- $\bar{e} = \bar{e}_p$ ones vector; $\bar{E} = \bar{E}_k = \bar{e}_k \bar{e}_k^T$ ones matrix

Model details

global minimum-energy conformation (GMEC)

Choose one rotamer from each set $\mathcal{V}_i$; minimize sum of weights/energies on edges in E.

- $m := (m_1 \quad \dots \quad m_p)^T$ size of subsets $\mathcal{V}_i$.
- $n_0 = |\mathcal{V}| \, (= \sum_k m_k)$
- $n := n_0 + 1$ size of matrices in SDP relaxation.

Quadratic integer programming (QIP) model

Computing the GMEC

$$
\begin{aligned}
& & val_{QIP} = \min_{x} & \sum_{(u,v)\in\mathcal{E}} E_{uv} x_u x_v & \\
& \text{(QIP)} & \text{s.t.} & \sum_{u\in\mathcal{V}_k} x_u = 1, & \forall\, k = 1,\ldots,p, \\
& & & x_u \in \{0,1\}, & \forall u \in \mathcal{V}\,,
\end{aligned}
$$

$$
x_u = \begin{cases} 1 & \text{if rotamer } u \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}
$$

Prepare model for lifting

Change to quadratic; Lift and Relax

Let $x := (x_u)$ and $y = \begin{pmatrix} 1 \\ x \end{pmatrix}$.

Lift to symmetric matrix space with

$$Y = yy^T, \qquad (\succeq 0)$$

i.e., Y_{uv} represents product $x_u x_v$, Y_{0v} represents $1 x_v$

Relax: ignore the (hard) rank one constraint on Y.

Zero-one variables

Change to quadratic $x_u^2 - x_u = 0$ This translates to the arrow constraint in the lifting: for Y (row-0 equals diagonal)

Efficiency versus strength of relaxation

Few constraints or many?

Few constraints means fewer constraints in the SDP relaxation. But adding more redundant constraints in the model means a possibly strenghtened SDP relaxation.

SDP is the Dual of Lagrangian relaxation

- Minimizing a quadratic subject to quadratic constraints leads to a Lagrangian dual which is the $\max_\lambda \min_x L(x, \lambda)$, where L is quadratic in x.
  (Thus more constraints implies stronger relaxation.)
- This leads to the constraint that the Hessian of the Lagrangian is positive semidefinite, an SDP.
- Take dual again; yields SDP relaxation of the original problem.

Matrix formulation for QIP

relabel the n_0 nodes in $\mathcal{V}$

$\mathcal{V}_1 \cong \{1, \ldots, m_1\}, \mathcal{V}_2 \cong \{m_1 + 1, \ldots, m_1 + m_2\}, \ldots,$
$\mathcal{V}_p \cong \left\{\left(\sum_{k=1}^{p-1} m_k\right) + 1, \ldots, n_0\right\}.$ and

complete definition $E_{uv} = 0$ if $(u, v) \notin \mathcal{E}$

define assignment type matrix $A \in \{0, 1\}^{p \times n_0}$

$$A := \begin{bmatrix} \bar{e}_{m_1}^T & 0 & 0 & \cdots & 0 \\ 0 & \bar{e}_{m_2}^T & 0 & \cdots & 0 \\ 0 & 0 & \bar{e}_{m_3}^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \bar{e}_{m_p}^T \end{bmatrix}; \quad A^T A = \begin{bmatrix} \bar{E}_{m_1} & 0 & 0 & \cdots & 0 \\ 0 & \bar{E}_{m_2} & 0 & \cdots & 0 \\ 0 & 0 & \bar{E}_{m_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \bar{E}_{m_p} \end{bmatrix},$$

QIP in matrix notation

Using A

$$
\text{(QIP)} \qquad
\begin{aligned}
val_{QIP} = \min_{x} \quad & x^T E x \\
\text{s.t.} \quad & Ax - \bar{e}_p = 0 \in \mathbb{R}^p \\
& x = \begin{bmatrix} v_1^T & v_2^T & \cdots & v_p^T \end{bmatrix}^T \in \{0,1\}^{n_0} \\
& v_k \in \{0,1\}^{m_k}, \ \forall\, k = 1, \ldots, p.
\end{aligned}
$$

QIP as QQP and redundant constraints within {}

$$
\text{(QQP)} \qquad
\begin{aligned}
val_{QIP} = val_{QQP} = \min_{x} \quad & x^T E x \\
\text{s.t.} \quad & \|\bar{e}_p - Ax\|^2 = 0 \\
& x \circ x - x = 0 \\
& \left\{ \begin{array}{l} (A^T A - I) \circ (xx^T) = 0 \\ (xx^T)_{ij} \geq 0, \ \forall\, (i,j) \in \mathcal{I}, \end{array} \right\}
\end{aligned}
$$

where: $\circ$ is Hadamard/elementwise product (forces zeros in Y) and $\mathcal{I} \subseteq \{(i,j) : 1 \leq i < j \leq n_0\}$ are valid inequalities

Forming the SDP relaxation

Start with QQP model with many constraints; apply recipe

1. form the Lagrangian relaxation;
2. apply homogenization;
3. simplify to obtain the dual and an equivalent SDP;
4. take the dual of dual to obtain the SDP relaxation of the original QIP
5. if strict feasibility fails, then apply facial reduction;
   -find the minimal face; obtain smaller problem with substitution $Y = W\bar{Y}W^T, \quad W \in \mathbb{R}^{n_0 \times t}, t < n_0$.
6. remove any redundant (linearly dependent) constraints.

Facial reduction as preprocessing

Exploit $Ax - \bar{e}_p = 0 \in \mathbb{R}^p$ constraint

Equivalently:

$$\begin{aligned} 0 &= e_i^T(Ax - \bar{e}_p), \quad \forall i = 1, \ldots, p \\ &= x^T A^T e_i - 1, \quad \forall i = 1, \ldots, p \\ &= \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} -1 \\ A^T e_i \end{pmatrix}, \quad \forall i = 1, \ldots, p \end{aligned}$$

Let $V = \begin{bmatrix} \begin{pmatrix} -1 \\ A^T e_1 \end{pmatrix} & \cdots & \begin{pmatrix} -1 \\ A^T e_p \end{pmatrix} \end{bmatrix}$. Then $y^T V = 0$. Therefore we can add the equivalent constraint to the SDP relaxation

$$Y(VV^T) = 0.$$

If range of W (full column rank) equals null space of V^T, then facial reduction (smaller $\bar{Y}$) is:

$$Y = W\bar{Y}W^T.$$

Form of SDP relaxation? ($\langle \cdot, \cdot \rangle$ trace inner prod.)

(DSDP-1)

$$
\begin{aligned}
d_{\mathcal{I}}^{**} := \quad \min_{Y} \quad & \left\langle \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix}, Y \right\rangle = \langle E, \bar{Y} \rangle \\
\text{s.t.} \quad & Y_{00} = 1 \\
& {}^{\mathrm{e}}\mathrm{bdiag}(Y) = p \\
& \mathrm{arrow}(Y) = 0 \\
& {}^{\mathrm{d}}\mathrm{bdiag}(Y) = 0 \\
& \mathcal{P}_{\mathcal{I}}(Y) \geq 0 \\
& Y = \begin{bmatrix} Y_{00} & y^T \\ y & \bar{Y} \end{bmatrix} \succeq 0.
\end{aligned}
$$

Gangster operator

shoots holes/zeros in the matrix Y; guarantees that the diagonal blocks are diagonal matrices.

Smaller primal-dual pair - satisfying strong p-d duality

$$
\begin{aligned}
d_{\mathcal{I}}^{**} = \min_{X} \quad & \left\langle \hat{E}, X \right\rangle \\
\text{s.t.} \quad & \text{arrow}(X) = 0, \\
& {}^{\text{d}}\text{bdiag}(X) = 0, \\
& X_{00} = 1, \\
& X \succeq 0,\ X \in \mathcal{S}^{n-p}, \\
& (WXW^T)_{ij} \geq 0, \quad \forall (i,j) \in \mathcal{I},
\end{aligned}
$$

and: $\hat{E} := W^T \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix} W$, $B_k := \begin{bmatrix} I_{k-1} \\ -\bar{e}_{k-1}^T \end{bmatrix} \in \mathbb{R}^{k \times (k-1)}$

$$
W = \begin{array}{c} \\ 1 \\ m_1 \\ m_2 \\ \\ m_p \end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & m_1-1 & m_2-1 & & m_p-1 \end{array} \\
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
e_{m_1} & B_{m_1} & 0 & \cdots & 0 \\
e_{m_2} & 0 & B_{m_2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
e_{m_p} & 0 & 0 & \cdots & B_{m_p}
\end{bmatrix}
\end{array},
$$

Dual SDP

$$
\begin{aligned}
d_{\mathcal{I}}^{**} = \quad & \max_{t,w,\Lambda,\xi} \quad t \\
& \text{s.t.} \quad {}^{1}\mathcal{O}(t) + \text{Arrow}(w) + {}^{\text{d}}\text{BDiag}(\Lambda) \\
& \qquad\qquad + \textstyle\sum_{(i,j)\in\mathcal{I}} \xi_{ij} W^T (e_i e_j^T + e_j e_i^T) W \preceq \hat{E} \\
& \xi \geq 0,\ \xi \in \mathbb{R}^{|\mathcal{I}|}.
\end{aligned}
$$

We have both primal and dual strong duality, i.e.,we have a zero duality gap and attainment.

Implementationn/Heuristics

Cutting planes

- start with small initial set $\mathcal{I} \subset \mathcal{I}_{\geq 0}$; corresponding to largest entries in E
- add most violated constraints, i.e., $Y_{ij} = (WXW^T)_{ij}$ is negative and $E_{ij}(WXW^T)_{ij}$ is very negative

Obtaining a good approximation for QIP from SDP

- Perron-Frobenius rounding: normalized eigenvector (largest) of Y^*: $u' := \frac{p}{u_2+\cdots+u_n}(u_2, \ldots, u_n) \in \mathbb{R}^{n_0}$ satisfies $Au' = \bar{e}_p$, and $u' \geq 0$ if $Y^* \geq 0$. (Empirically true even without nonnegativity.)
- Projection rounding: use diagonal $\begin{pmatrix} 1 \\ u'' \end{pmatrix}$ of the optimal solution Y^* is used. Again, u'' satisfies $Au'' = \bar{e}_p$, $u'' \geq 0$.

Data set/Numerics

Four Methods

1. original SDP relaxation
2. SDP and facial reduction
3. SDP and cutting planes
4. SDP and facial reduction and cutting planes

26 protein data from PDB of various sizes

`SCPCP` consistently produces

- shorter cpu time,
- higher accuracy of SDP solution, and
- importantly, better integer solutions from rounding (essentially optimal - close to dual optimal value)

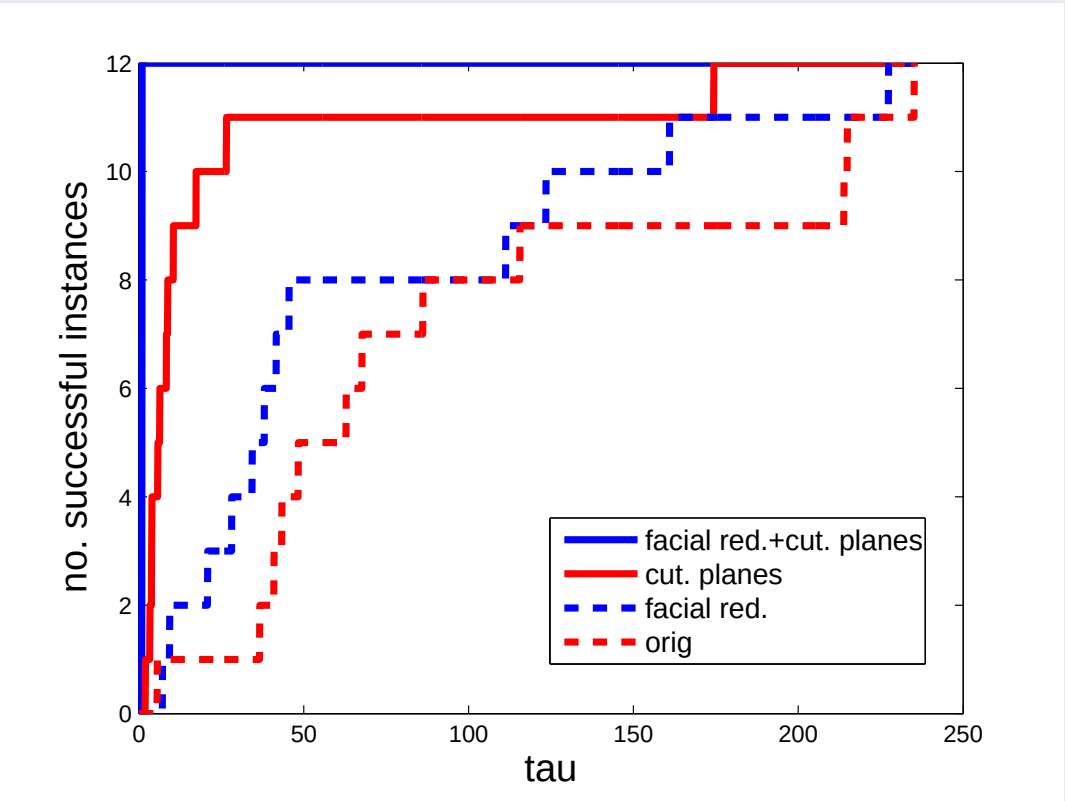
Performance Profile

$t_{i,j}$:= run time for QIP final solution, instance i method j

$1 \leq r_{i,j} := \frac{t_{i,j}}{\min\{t_{i,j}: j=1,2,3,4\}}$ perform. ratio method j on instance i

$\rho(\tau)$:= number of instances i such that $r_{i,j} \leq \tau$

Figure: Performance profile comparing the four methods



Medium sized triose phosphate isomerase, 1TIM

Table: Information on input data for 1TIM

Total number of residues / partitions	249
Total number of rotamers / nodes	819
Number of energy values / edges	66520
$\max_{i,j} E_{i,j}$	5.80e+15
$\min_{i,j} E_{i,j}$	-7.7783
Number of valid nonnegativity constraints $\left(= \frac{1}{2}\left(n_0^2 - \sum_{k=1}^{p} m_k^2\right)\right)$	329760

Table: Information on output for 1TIM

Increments in cuts	**100**	**120**	**180**
Total time elapsed (hr)	2.51	2.16	1.36
Number of iterations	12	11	9
Final number of nonneg. constr.	2306	2247	2217
Percentage of valid nonneg. constr. used	0.70 %	0.68 %	0.67%
dual SDP optval	685.61	685.61	685.61
objval for QIP	685.61	685.61	685.61
relative diff	5.81e-12	8.68e-12	4.62e-13

Two tables

Table 3 Results on small proteins

Protein	n_0	p	run time (sec)		dual SDP optval		objval in IQP		relative diff		relative gap	
			SCPCP	[6]	SCPCP	[6]	SCPCP	[6]	SCPCP	[6]	SCPCP	[6]
1AAC	117	85	6.58	296.06	-206.33	-206.33	-206.33	-206.33	5.75E-11	1.72E-05	1.30E-09	4.21E-04
1AHO	108	54	7.97	364.73	33.53	33.53	33.53	33.53	8.44E-11	4.95E-05	2.45E-09	4.68E-04
1BRF	130	45	14.96	977.08	-31.11	-31.11	-31.11	-31.11	3.92E-11	2.27E-05	3.08E-09	1.24E-04
1CC7	160	66	28.60	1059.06	-63.76	-2.30E+07	-63.76	3.73E+04	1.13E-11	2.01	1.27E-09	1.11
1CKU	115	60	5.46	815.18	113.83	113.83	113.83	113.83	7.17E-11	4.79E-05	3.42E-09	1.13E-04
1CRN	65	37	12.76	46.42	-14.87	-14.87	-14.87	-14.87	1.64E-12	3.05E-05	2.20E-10	3.66E-04
1CTJ	153	61	16.15	777.31	-129.53	-6.69E+06	-129.53	174.65	2.98E-11	2.00	2.29E-09	1.07
1D4T	188	89	41.32	2775.34	-173.03	-2.96E+07	-173.03	291.13	3.88E-11	2.00	1.35E-09	1.20
1IGD	82	50	5.51	189.04	-69.25	-69.25	-69.25	-69.25	4.79E-10	2.74E-06	5.76E-09	3.39E-05
1PLC	129	82	14.32	1766.03	-1.50	-1.50	-1.50	-1.50	1.28E-11	7.28E-04	4.60E-10	1.09E-03
1VFY	134	63	23.49	1765.36	-90.09	-90.09	-90.09	-90.09	1.67E-11	-1.11E-05	9.15E-10	3.79E-05
4RXN	98	48	18.44	366.48	-21.65	-21.65	-21.65	-21.65	1.48E-11	2.62E-05	4.19E-10	6.67E-05

Table 4 Results on medium-sized proteins

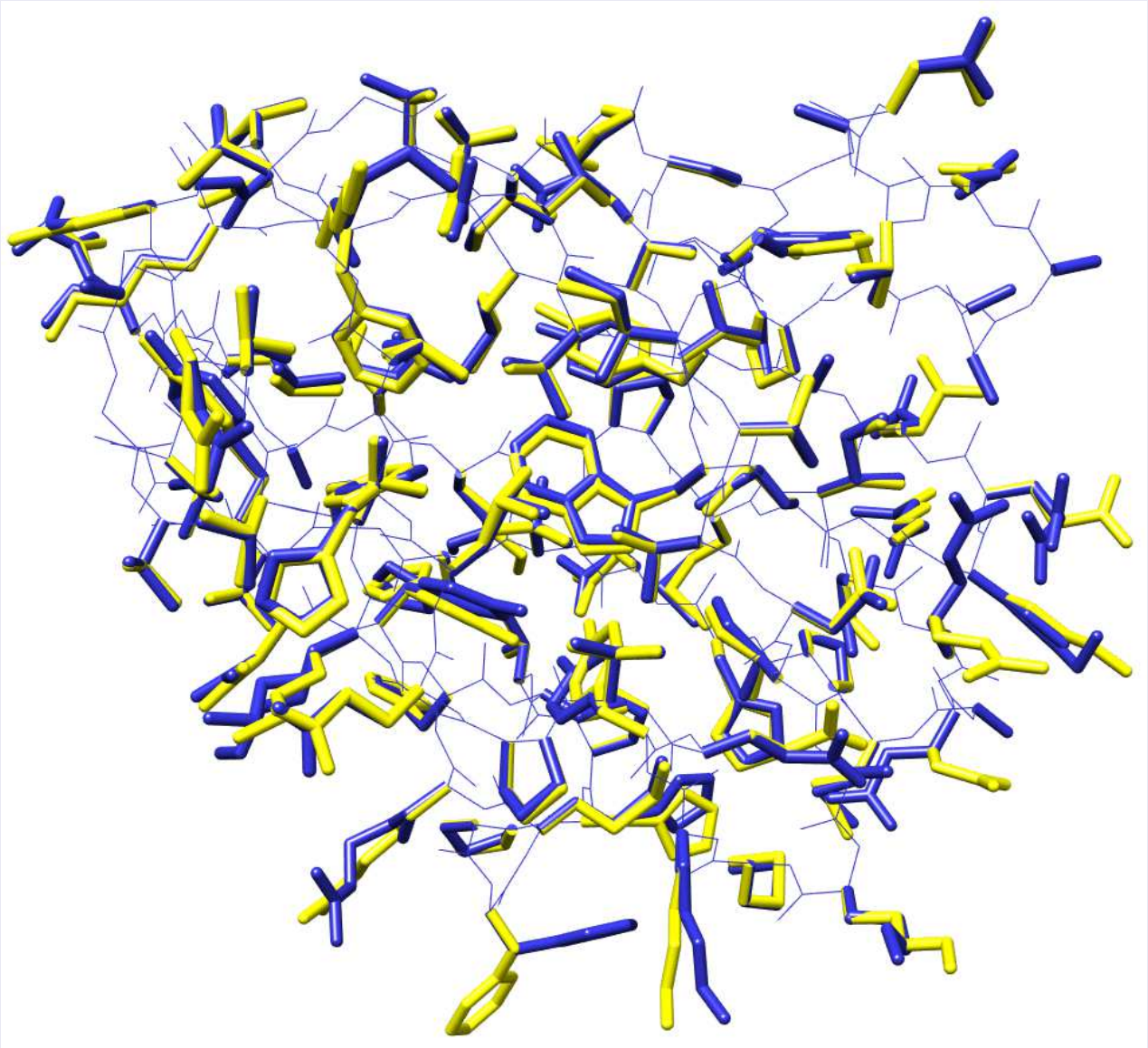
Protein	n_0	p	run time (min)		dual SDP optval		objval in IQP		relative diff		relative gap	
			SCPCP	[6]	SCPCP	[6]	SCPCP	[6]	SCPCP	[6]	SCPCP	[6]
1B9O	265	112	0.64	254.85	-140.24	-5.63E+07	-140.24	1.91E+06	1.19E-11	2.14	1.45E-09	1.24
1C5E	200	71	2.59	70.63	-131.75	-6.46E+04	-131.75	148.82	4.93E-11	2.01	5.02E-09	1.00
1C9O	207	53	2.15	66.50	-83.55	-1.88E+06	-83.55	1628.10	3.35E-12	2.00	2.77E-10	1.02
1CZP	237	83	1.90	143.95	-37.88	-2.26E+04	-37.88	1254.42	8.30E-11	2.24	1.03E-08	1.00
1MFM	216	118	0.19	102.11	-201.29	-7.36E+07	-201.29	1369.92	2.01E-11	2.00	1.24E-09	1.09
1QQ4	365	143	5.70	-	-102.40	-	-102.40	-	6.49E-11	-	2.27E-08	-
1QTN	302	134	5.04	-	-178.77	-	-178.77	-	2.24E-11	-	4.12E-09	-
1QU9	287	101	7.55	-	-124.96	-	-124.96	-	1.80E-11	-	5.52E-09	-

Large Scale Case

Table 5 Results on large proteins (SCPCP only)

Protein	n_0	p	run time (hr)	dual SDP optval	Objval in IQP	rel. diff	rel. gap	numcut	# iter	Final # cuts
1CEX	435	146	0.08	140.20	140.20	1.26E-11	5.57E-09	40	9	485
1CZ9	615	111	3.96	497.46	497.46	2.98E-13	6.37E-10	60	25	1997
1QJ4	545	221	0.15	-286.83	-286.83	5.31E-12	1.14E-09	60	14	1027
1RCF	581	142	0.85	-191.54	-191.54	3.71E-12	1.15E-08	60	17	1305
2PTH	930	151	29.65	-159.41	-159.41	8.69E-09	7.63E-06	120	34	7247
5P21	464	144	0.31	-135.75	-135.75	1.39E-12	7.33E-10	40	16	822

Figure: Superposition of the reconstruction (light grey) of 1AAC over the crystallized form described in the PDB (dark grey)



Summary

- We model protein design using using a QIP and transform to a quadratic-quadratic model
- Lagrangian Relaxation leads to an SDP program and the dual is the SDP relaxation
- Adding redundant constraints strengthens the SDP relaxation
- The strict feasibility fails for SDP relaxation; but, it can be exploited using facial reduction to get a smaller/stable problem
- Cutting planes help yield stronger approximate solutions. Empirical evidence shows efficiency and robustness of adding redundant constraints and applying facial reduction.

Thanks for your attention!

Facial reduction for semidefinite programming and its application for the selection of rotamers in protein conformations

Henry Wolkowicz

Dept. Combinatorics and Optimization,
University of Waterloo

with: Forbes Burkowski and Yuen-Lam (Cheung) Voronin