

Supplement to ECCB 2010: A Convex Programming Model for Protein Structure Prediction

Chris Kauffman and George Karypis

April 12, 2010

1 Constraint Violations

The RMSD between predicted and true structure does not always tell the full story of the prediction quality. It is also instructive to examine how badly the predicted structure violates the equality and inequality distance constraints. Recall these constraints are encoded by the equation

$$A_l x_l + A_s x_s = b$$

for vectorized Gram matrix x_s and nonnegative distance slacks x_l . For estimated three-dimensional coordinates $\hat{R}$ we use the corresponding Gram matrix $\hat{G} = \hat{R}^T \hat{R}$ which is vectorized to $\hat{g}$. The distance slacks for $\hat{g}$ are called $\hat{g}_l$ and are used in the calculation of the inequality constraints. Should an inequality constraint be violated, the corresponding distance slack is set to zero. The level of constraint violations ν is then

$$\nu = \frac{\|A_l \hat{g}_l + A_s \hat{g} - b\|_2}{\|b\|_2}, \quad (1)$$

which is zero when no violations occur.

Intuitively, ν combines two measures of badness for the structures. First, deviations of backbone bond distances from their assumed known values will add to ν . Second, atom clashes in which two nonbonded atoms are too close to one another contribute to ν .

Figure 1 shows the level of distance constraint violation for the distance geometry methods at several noise levels. The quality of CMDS according to mean RMSD is tempered by its inability to avoid atom clashes. This is illustrated by increasing levels of constraint violation as the noise level increases. NEDM and Domo satisfy the constraints much more readily with Domo doing so slightly better than NEDM.

Figure 2 shows the constraint violations of the pure objective based methods. These methods are competitive with Domo and NEDM in their ability to produce structures that satisfy constraints.

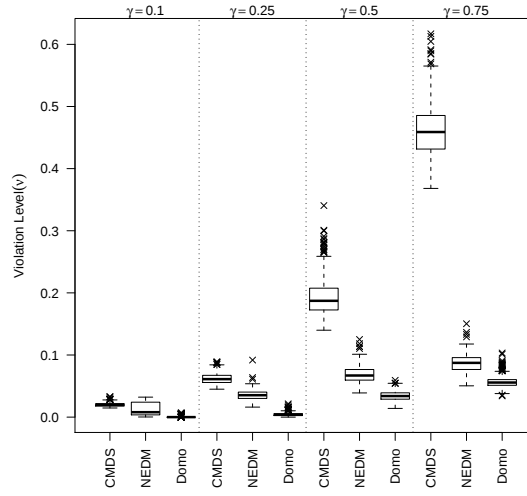


Figure 1: Box and whisker plot of the level of constraint violation for the distance geometry methods at four noise levels.

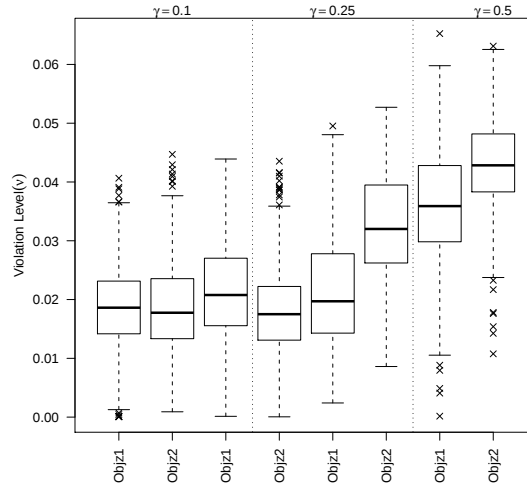


Figure 2: Box and whisker plot of the level of constraint violation for the pure convex prediction methods at four noise levels.