David J. Elkind - LIPO

I'm interested in estimating the Lipschitz constant k for functions which approximate as Lipschitz. When the functions are noisy, they can't be Lipschitz, but adding a noise term s can make the functions behave more nicely; including the large penalty 10^6 for s forces it to often be 0.

The function of interest is f and it is only observed via some finite number of points $(x_i, f(x_i))$ for $i = 1, 2, \ldots t$. Define

$$U(y) = \min_{i \in 1,2,\ldots t} \left[f(x_i) + \sqrt{s_i + (y - x_i)^\top K (y - x_i)} \right] \tag{1}$$

One way to estimate k is to optimize the following program (using a different k_i for each coordinate):

$$\begin{aligned}
\min_{K,s} \quad & ||K||_F^2 + 10^6 \sum_{i=1}^{t} s_i^2 && (2)\\
\text{s.t.} \quad & U(x_i) \geq f(x_i) i \in \{1, 2, \ldots t\} && (3)\\
& s_i \geq 0 \forall i \in \{1, 2, \ldots t\} && (4)\\
& K_{i,j} \geq 0 \forall i, j \in \{1, 2, \ldots d\}^2 && (5)\\
& \text{K is a diagonal matrix.} && (6)
\end{aligned}$$

Expressing a signle constraint, we have

$$\begin{aligned}
U(x_i) &\geq f(x_i) \ \forall i = 1, 2, \ldots, t && (7)\\
\implies f(x_j) + \sqrt{s_j + (x_i - x_j)^\top K (x_i - X_j)} &\geq f(x_i) && (8)\\
s_j + (x_i - x_j)^\top K (x_i - x_j) &\geq (f(x_i) - f(x_j))^2 && (9)\\
& && (10)
\end{aligned}$$

Which is linear in s and K. Moreover, this is exactly the definition of Lipschitz continuity when $s_j = 0$.

This is a little bit of a bummer, though, since this isn't symmetric in i, j due to the presence of the s_j. This implies that there are $t^2 > \binom{t}{2}$ constraints, which might be too many for large problems.