For a given dimension d, number of sets s and a series of measurements $M_i^k \in \mathbb{C}^d \otimes \mathbb{C}^d$ our problem is as follows:

$$\min -\sum_{i<j}^{s}\sum_{k,l}^{d}\sqrt{1-\operatorname{tr}(M_k^i M_l^j)} \tag{1}$$

subject to:

$$M_k^i \succeq 0 \tag{2}$$
$$\sum_k^d M_k^i = \mathbb{I} \tag{3}$$
$$\operatorname{tr}(M_k^i) = 1 \tag{4}$$
$$(M_k^i)^2 = M_k^i \tag{5}$$

Need to define three functions for this method: the objective function $f(x)$, the constraint function $g(x)$ and the function $X(x)$ converting between the vector form x and the matrix form such that $X(x)$ is positive semidefinite. These functions do not need to be linear.

In our case $X(x)$ takes the form:

$$X(x) = \sum_a^n x_a A_a + B \tag{6}$$

such that it turns a vector $x \in \mathbb{R}^n$ containing the unique elements of the measurements into a matrix $X(x) \in \mathbb{R}^p \otimes \mathbb{R}^p$ which contains the real components on the block diagonals and the imaginary components on the off-diagonals, with $p = 2d^2n$. For instance, in the $d = 2$, $n = 2$ case, if each $M_i^k = R_i^k + iI_i^k$:

$$X(x) = \begin{pmatrix} R_1^1 & 0 & 0 & 0 & I_1^1 & 0 & 0 & 0 \\ 0 & R_2^1 & 0 & 0 & 0 & I_2^1 & 0 & 0 \\ 0 & 0 & R_1^2 & 0 & 0 & 0 & I_1^2 & 0 \\ 0 & 0 & 0 & R_2^2 & 0 & 0 & 0 & I_2^2 \\ I_1^1 & 0 & 0 & 0 & R_1^1 & 0 & 0 & 0 \\ 0 & I_2^1 & 0 & 0 & 0 & R_2^1 & 0 & 0 \\ 0 & 0 & I_1^2 & 0 & 0 & 0 & R_1^2 & 0 \\ 0 & 0 & 0 & I_2^2 & 0 & 0 & 0 & R_2^2 \end{pmatrix} \tag{7}$$

Many of these elements are defined in relation to others such that the vector x contains as little information as needed, whilst also forcing the submatrices of any $X(x)$ to satisfy the identity and trace constraints. This also has the benefit of meaning in our case $g(x)$ is only used for the projective constraint.

Splitting the objective function into real and imaginary components and using the identity $\mathrm{tr}(M_k^i M_l^j) = M_k^i \cdot M_l^j$:

$$f(x) = -\sum_{i<j}^{s}\sum_{k,l}^{d}\sqrt{1 - R_k^i \cdot R_l^j + I_k^i \cdot I_l^j} \tag{8}$$

Now defining the extraction matrices C_k^i, D_k^i, E_k^i and F_k^i such that $C_k^i X D_k^i = R_k^i$ and $E_k^i X F_k^i = I_k^i$ with the notation that $X = X(x)$:

$$f(x) = -\sum_{i<j}^{s}\sum_{k,l}^{d}\sqrt{1 - C_k^i X D_k^i \cdot C_l^j X D_l^j + E_k^i X F_k^i \cdot E_l^j X F_l^j} \tag{9}$$

Now taking the first derivative of this, letting d_{kl}^{ij} be the value inside the above square root:

$$\begin{aligned}
\frac{\partial f(x)}{\partial x_b} = -\sum_{i<j}^{s}\sum_{k,l}^{d}\frac{1}{2}(d_{kl}^{ij})^{-\frac{1}{2}}(&-C_k^i A_b D_k^i \cdot C_l^j X D_l^j - C_k^i X D_k^i \cdot C_l^j A_b D_l^j &\quad (10)\\
&+E_k^i A_b F_k^i \cdot E_l^j X F_l^j + E_k^i X F_k^i \cdot E_l^j A_b F_l^j) &\quad (11)
\end{aligned}$$

Then the second derivative, remembering that d_k^i has a dependence on x:

$$\begin{aligned}
\frac{\partial^2 f(x)}{\partial x_b \partial x_c} = \sum_{i<j}^{s}\sum_{k,l}^{d}\frac{1}{4}(d_{kl}^{ij})^{-\frac{3}{2}}(&-C_k^i A_b D_k^i \cdot C_l^j X D_l^j - C_k^i X D_k^i \cdot C_l^j A_b D_l^j &\quad (12)\\
&+E_k^i A_b F_k^i \cdot E_l^j X F_l^j + E_k^i X F_k^i \cdot E_l^j A_b F_l^j) &\quad (13)\\
&(-C_k^i A_c D_k^i \cdot C_l^j X D_l^j - C_k^i X D_k^i \cdot C_l^j A_c D_l^j &\quad (14)\\
&+E_k^i A_c F_k^i \cdot E_l^j X F_l^j + E_k^i X F_k^i \cdot E_l^j A_c F_l^j) &\quad (15)\\
-\frac{1}{2}(d_{kl}^{ij})^{-\frac{1}{2}}(&-C_k^i A_b D_k^i \cdot C_l^j A_c D_l^j - C_k^i A_c D_k^i \cdot C_l^j A_b D_l^j &\quad (16)\\
&+E_k^i A_b F_k^i \cdot E_l^j A_c F_l^j + E_k^i A_c F_k^i \cdot E_l^j A_b F_l^j) &\quad (17)
\end{aligned}$$

Computationally many things here are different, such that X is only ever calculated once per iteration and the extraction matrices are unneeded, instead the submatrices are extracted directly from the cached X using Eigen's "X.block()" routine, which claims $O(0)$ scaling when compiled with optimisations. Note that in the above expressions many of these terms are zero, for instance $C^i_k A_b D^i_k \cdot C^j_l A_b D^j_l = 0$ since each A matrix will only have non-zero components for a single measurement, so one of those two extractions must result in a zero matrix.

In order to enforce that the measurements are projectors we define the constraint function $g(x)$:

$$
\begin{aligned}
g(x) &= |X^2 - X|^2 && (18)\\
&= (X^2 - X) \cdot (X^2 - X) && (19)
\end{aligned}
$$

And its first derivative:

$$
\begin{aligned}
\frac{\partial g(x)}{\partial x_b} &= 2(X^2 - X) \cdot \frac{\partial}{\partial x_b}(X^2 - X) && (20)\\
&= 2(X^2 - X) \cdot (2A_b X - A_b) && (21)
\end{aligned}
$$

And its second derivative:

$$
\begin{aligned}
\frac{\partial^2 g(x)}{\partial x_b \partial x_c} &= \frac{\partial}{\partial x_c} 2(X^2 - X) \cdot (2A_b X - A_b) && (22)\\
&= \left[\frac{\partial}{\partial x_c} 2(X^2 - X)\right] \cdot (2A_b X - A_b) + 2(X^2 - X) \cdot \left[\frac{\partial}{\partial x_c}(2A_b X - A_b)\right] && (23)\\
&= [2(2A_c X - A_c)] \cdot (2A_b X - A_b) + 2(X^2 - X) \cdot [2A_b A_c] && (24)\\
&= 2\ (2A_c X - A_c) \cdot (2A_b X - A_b) + 4(X^2 - X) \cdot A_b A_c && (25)
\end{aligned}
$$

Putting these all together we can construct the Lagrangian and it's first/second derivatives, which are then used as in the paper.

$$
L(x) = f(x) - y g(x) - X \cdot Z \qquad (26)
$$

$$
\nabla L(x) = \nabla f(x) - y \nabla g(x) - \nabla (X \cdot Z) \qquad (27)
$$

Now for a KKT point:

$$\nabla L(x) = 0 \qquad \text{and} \qquad g(x) = 0 \qquad \text{and} \qquad Z = 0 \tag{28}$$
$$\implies \nabla f(x) = 0 \tag{29}$$

$$\text{for} \qquad b = 1, ..., n \tag{30}$$
$$\sum_{i<j}^{s}\sum_{k,l}^{d}(d_{kl}^{ij})^{-\frac{1}{2}}(-C_k^i A_b D_k^i \cdot C_l^j X D_l^j - C_k^i X D_k^i \cdot C_l^j A_b D_l^j \tag{31}$$
$$+E_k^i A_b F_k^i \cdot E_l^j X F_l^j + E_k^i X F_k^i \cdot E_l^j A_b F_l^j) = 0 \tag{32}$$
$$\implies \sum_{i<j}^{s}\sum_{k,l}^{d}(d_{kl}^{ij})^{-\frac{1}{2}}(-C_k^i A_1 D_k^i \cdot C_l^j X D_l^j - C_k^i X D_k^i \cdot C_l^j A_1 D_l^j \tag{33}$$
$$+E_k^i A_1 F_k^i \cdot E_l^j X F_l^j + E_k^i X F_k^i \cdot E_l^j A_1 F_l^j) = 0 \tag{34}$$

This A_1 only has non-zero components in the real section and for $i = 1$:

$$\sum_{j\neq 1}^{s}\sum_{k,l}^{d}(d_{kl}^{1j})^{-\frac{1}{2}} C_k^1 A_1 D_k^1 \cdot C_l^j X D_l^j = 0 \tag{35}$$

This A_1 is also only non-zero for $k = 1$ and $k = d$:

$$\sum_{j\neq 1}^{s}\sum_{l}^{d}[(d_{1l}^{1j})^{-\frac{1}{2}} C_1^1 A_1 D_1^1 \cdot C_l^j X D_l^j + (d_{dl}^{1j})^{-\frac{1}{2}} C_d^1 A_1 D_d^1 \cdot C_l^j X D_l^j] = 0 \tag{36}$$

Moving things as far out of the sum as we can and using that $C_1^1 A_1 D_1^1 = -C_d^1 A_1 D_d^1$:

$$C_1^1 A_1 D_1^1 \cdot \sum_{j\neq 1}^{s}\sum_{l}^{d} \; [(d_{1l}^{1j})^{-\frac{1}{2}} - (d_{dl}^{1j})^{-\frac{1}{2}}] \; C_l^j X D_l^j = 0 \tag{37}$$

Since this can be repeated for all the A matrices referring to the real parts of measurement matrix 1, each of which are linearly independent, all of the components of the sum on the right must equal zero:

$$\sum_{j\neq 1}^{s}\sum_{l}^{d}\ [(d_{1l}^{1j})^{-\frac{1}{2}}-(d_{dl}^{1j})^{-\frac{1}{2}}]\ C_l^j X D_l^j = 0 \tag{38}$$

Repeating for other matrices, real and imaginary components:

$$\text{for} \qquad i\in 1,...,s \qquad k\in 1,...,d-1 \tag{39}$$

$$\sum_{j\neq i}^{s}\sum_{l}^{d}\ [(d_{kl}^{ij})^{-\frac{1}{2}}-(d_{dl}^{ij})^{-\frac{1}{2}}]\ C_l^j X D_l^j = 0 \tag{40}$$

$$\sum_{j\neq i}^{s}\sum_{l}^{d}\ [(d_{kl}^{ij})^{-\frac{1}{2}}-(d_{dl}^{ij})^{-\frac{1}{2}}]\ E_l^j X F_l^j = 0 \tag{41}$$

For non-zero X (the non-trivial solution) this implies the d's are all equal and thus the moments are equal. Also requiring that $g(x)=0$ means the matrices have to be projective so therefore the only solution is when each $d_{kl}^{ij}=1-1/d$, thus $\nabla f(x)=0 \implies$ MUBs. This also means that there exist no local minima, however the search space is still convex since it consists of many disconnected convex regions which all have the same local/global minima which are all stationary points, assuming MUBs. In the no MUB case there are no stationary points but the local minima also seem to be the global.