The set spanned by the affine hyperplane is given by

$$\mathcal{A} = \{x \in \mathbb{R}^3 : n^{\mathrm{T}}(x - \beta_1) = 0\} \tag{1}$$

where n is unit normal vector and β_1 is the target pose translation. The normal vector n is found using the head rotation matrix H (that is constructed from yaw, pitch and roll angles) and the unit z-axis vector.

With reference to fig. **??** we note that transforming the the pose and vector measurements in the gazer's coordinate system into the target's coordinate system is given by

$$P_2^{C1} = RP_2^{C2} + d \tag{2}$$

and

$$v_2^{C1} = Rv_2^{C2} \tag{3}$$

where R (orthonormal rotation matrix) and d (distance vector) are the two quantities we are trying to solve for. We use the $C1$ in P_2^{C1} to denote that the coordinate is in camera one's coordinate system. Then the projection of a gaze vector is given by

$$\hat{P_2^{C1}} = \alpha v_2^{C1} + P_2^{C1} \tag{4}$$

where α is the fluctuating distance between the two speakers given by $\alpha = \|s\|_2$ of the line s shown in fig. **??**. Given estimates of R and d, and substituting eq. 4 into the hyperplane equation eq. 1, the value for α that causes $\hat{P_2^{C1}}$ to intersect with the hyperplane is given by

$$\alpha = \frac{n^{\mathrm{T}} P_1^{C1} - n^{\mathrm{T}} \hat{P_2^{C1}}}{n^{\mathrm{T}} v_2^{C1}} \tag{5}$$

The vector that corresponds to the the x,y projections on the hyperplane is then given by

$$y = HP_2^{C1} - HP_1^{C1} + \alpha(Hv_2^{C1}) \tag{6}$$

The z-element of this y vector is discarded as it will always be zero and the x,y elements are kept as the projection coordinates in $\mathbb{R}^2$.

The eye locations are projected onto an affine hyperplane given by the pose and head rotation information. The problem is then formulated as a non-linear least squares problem:

$$\hat{X} = \arg\min_X \frac{1}{2}\|F(X)\|^2 \tag{7}$$

where the distance we seek to minimize is the column vector:

$$F(X) = [x - y; \alpha - 1000] \tag{8}$$

Here we only show 3 elements (2 for y, and 1 for alpha). In practice we optimize for the pair of eyes and for both target/gazer configurations at the same time. This results in an optimization on 12 values. To find the inverse rotation and distance values for the swapped target/gazer configuration, we use the inverse of the homogeneous rotation/translation matrix.