Orthogonal Procrustes Problem

Consider we have a set of points $\boldsymbol{P}_1 \in \mathbb{R}^{2\times N}$ and another set of points $\boldsymbol{P}_2 \in \mathbb{R}^{2\times N}$ such that $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are related by an orthonormal transformation $\boldsymbol{R}$ such that $\boldsymbol{P}_1 = \boldsymbol{R}\boldsymbol{P}_2 + \boldsymbol{E}$ where $\boldsymbol{E} \in \mathbb{R}^{2\times N}$ is an error (or noise) matrix. The aim is to find $\boldsymbol{R}$ given $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$.
The standard least squares solution given by $\boldsymbol{R} = \boldsymbol{P_1}\boldsymbol{P_2}^T(\boldsymbol{P_2}\boldsymbol{P_2}^T)^{-1}$ will fail, because it will not impose the constraint that $\boldsymbol{R}$ is orthonormal (excepting the situation where $\boldsymbol{E} = \boldsymbol{0}$). You can try this out in MATLAB! To solve for $\boldsymbol{R}$, we seek to minimize the following quantity:

$$
\begin{aligned}
E(\boldsymbol{R}) &= \|\boldsymbol{P_1} - \boldsymbol{R}\boldsymbol{P_2}\|_F^2 &(1)\\
&= \text{trace}((\boldsymbol{P_1} - \boldsymbol{R}\boldsymbol{P_2})^T(\boldsymbol{P_1} - \boldsymbol{R}\boldsymbol{P_2})) &(2)\\
&= \text{trace}(\boldsymbol{P_1^T}\boldsymbol{P_1} + \boldsymbol{P_2^T}\boldsymbol{R^T}\boldsymbol{R}\boldsymbol{P_2} - \boldsymbol{P_2^T}\boldsymbol{R^T}\boldsymbol{P_1} - \boldsymbol{P_1^T}\boldsymbol{R}\boldsymbol{P_2}) &(3)\\
&= \text{trace}(\boldsymbol{P_1^T}\boldsymbol{P_1} + \boldsymbol{P_2^T}\boldsymbol{P_2} - \boldsymbol{P_2^T}\boldsymbol{R^T}\boldsymbol{P_1} - \boldsymbol{P_1^T}\boldsymbol{R}\boldsymbol{P_2}) &(4)\\
&= \text{trace}(\boldsymbol{P_1^T}\boldsymbol{P_1} + \boldsymbol{P_2^T}\boldsymbol{P_2}) - 2\text{trace}(\boldsymbol{P_1^T}\boldsymbol{R}\boldsymbol{P_2}) &(5)
\end{aligned}
$$

The second-last equality follows as $\boldsymbol{R^T}\boldsymbol{R} = \boldsymbol{I}$. The last equality follows using the fact that trace of a matrix equals the trace of its transpose.

The first two terms are constant and the last term has a negative sign. Hence, minimizing $E(\boldsymbol{R})$ w.r.t. $\boldsymbol{R}$ is equivalent to maximizing $\text{trace}(\boldsymbol{P_1^T}\boldsymbol{R}\boldsymbol{P_2})$ w.r.t. $\boldsymbol{R}$. Now, we have

$$
\begin{aligned}
\text{trace}(\boldsymbol{P_1^T}\boldsymbol{R}\boldsymbol{P_2}) &= \text{trace}(\boldsymbol{R}\boldsymbol{P_2}\boldsymbol{P_1^T})(\because \text{trace}(AB) = \text{trace}(BA)) &(6)\\
&= \text{trace}(\boldsymbol{R}\boldsymbol{U'}\boldsymbol{S'}\boldsymbol{V'^T}) \text{ using SVD of } \boldsymbol{P_2}\boldsymbol{P_1^T} = \boldsymbol{U'}\boldsymbol{S'}\boldsymbol{V'^T} &(7)\\
&= \text{trace}(\boldsymbol{S'}\boldsymbol{V'^T}\boldsymbol{R}\boldsymbol{U'}) = \text{trace}(\boldsymbol{S'}\boldsymbol{X}) \text{ where } \boldsymbol{X} \text{ is orthonormal} &(8)\\
&= \sum_i S'_{ii}X_{ii} &(9)
\end{aligned}
$$

The values S'_{ii} are all non-negative, and the above expression is maximized if $\boldsymbol{X}_{ii} = 1$ all along its diagonal. As $\boldsymbol{X}$ is orthonormal, we must have $\boldsymbol{X} = \boldsymbol{I}$, and hence $\boldsymbol{V'^T}\boldsymbol{R}\boldsymbol{U'} = \boldsymbol{I}$ giving $\boldsymbol{R} = \boldsymbol{V'}\boldsymbol{U'}^T$ where $\boldsymbol{U'}$ and $\boldsymbol{V'}$ are obtained from the SVD of $\boldsymbol{P_2}\boldsymbol{P_1^T}$.

But we are still not done - why? Because we found out a matrix that is guaranteed to produce the least error amongst all orthonormal matrices. But this orthonormal matrix may not be a rotation matrix, i.e. it may have determinant -1 instead of +1. This change of sign can become an issue if the amount of noise is very high. So what do we do? If the determinant of $\boldsymbol{R}$ is -1, replace it by $\boldsymbol{R}_{\text{final}} = \boldsymbol{V'}\boldsymbol{J}\boldsymbol{U'}^T$ where $\boldsymbol{J} = \text{diag}(1, -1)$ is a 2×2 diagonal matrix that negates the column of $\boldsymbol{V'}$ corresponding to the smaller singular value. What's the logic behind this? We know that $\sum_i S'_{ii}X_{ii} = S'_{11}X_{11} + S'_{22}X_{22}$ (it is a 2 by 2 matrix!). The maximum value is $S_{11} + S_{22}$ which gave us a reflection matrix. Now S_{11} and S_{22} are singular values and hence always non-negative. Also the values along the diagonal of $\boldsymbol{X}$ always lie in the range $[-1, 1]$ since $\boldsymbol{X}$ is orthonormal. Hence, the next highest value will be $S_{11} - S_{22}$ since the function is linear and can have extrema only at the following values of (X_{11}, X_{22}): $(-1, 1), (1, 1), (1, -1), (-1, -1)$. (This last argument is admittedly non-rigorous as the constraint $\det(\boldsymbol{R}) = 1$ is certainly non-linear. The answer however is indeed correct, but the fully rigorous explanation for it is rather long-winded and hairy. You can find it in a paper 'Least squares estimation of transformation parameters between two point patterns' by Shinji Umeyama, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991. For a student who is mathematically inclined and sufficiently patient, this is an excellent project topic.)

This is a deceptively tricky problem, as it is so very easy to fall into the 'pseudo-inverse' trap :-). However, it is an important and fundamental problem in computer vision, graphics and medical imaging (especially in a sub-branch called as 'statistical shape analysis'). It is popularly called as the 'orthogonal procrustes problem' and the first complete solution to it was first proposed by a person named Peter Schönemann in a psychology (!) journal in 1966. A popular application scene where this problem is useful is as follows: Consider N pairs of corresponding points marked out on two 3D models of a person (acquired from say a Kinect from two different positions). What is the rotation and translation in between these points? This has obvious applications in person recognition and aligning scans of humans or different types of objects.

Note that I have written this derivation for 2×2 rotation matrices, but it is applicable to 3D as well as other higher dimensions equally well.