We want to solve

$$\begin{bmatrix} \hat{\alpha}, \hat{\theta} \end{bmatrix} = \arg \min_{\alpha, \theta} \sum_{i=1}^{n} \| y_i - \hat{y}_i \|^2_C.$$ 

Here we assume that the objective function is

$$g(y_{io}) = \alpha + y_{io}^\top \theta = 0 \quad i = 1, \ldots, n$$

and we are in the homoscedastic case.

Like in the case of TLS, we assume that $\tilde{y}_i$ are the centered points of $y_i$. Similarly, $\tilde{Y}$ and $\tilde{Y}_o$ are the $n \times p$ matrices of the centered data and the centered true values.

The noise is

$$\delta y \sim GI(0, \sigma^2 C_y)$$

a i.i.d. symmetric distribution, full rank $p$ but not necessarily diagonal covariance.

We have $||\theta|| = 1$.

Will apply the Lagrange multipliers to solve the estimation

$$J_{GTLS}(\hat{\theta}, \hat{y}_1, \cdots, \hat{y}_n) = \frac{1}{2} \sum_{i=1}^{n} (\tilde{y}_i - \hat{y}_i)^\top C_y^{-1} (\tilde{y}_i - \hat{y}_i) + \sum_{i=1}^{n} \eta_i \hat{y}_i^\top \hat{\theta}$$

where each one of the $n$ factors (second term) has $\eta_i$ multiplied with a scalar zero since the variable estimate $\hat{y}_i$ satisfies the constraint. We take the gradient in the estimates values $\hat{y}_i$ equal to zero.

$$\frac{\partial J_{GTLS}}{\partial \hat{y}_i} = 0 \quad \hat{y}_i = \tilde{y}_i - \eta_i C_y \hat{\theta}$$

where we seek the $\hat{y}_i$ and $\hat{\theta}$. Multiplying both sides the $\hat{\theta}^\top$ we obtain

$$\hat{\theta}^\top \hat{y}_i = 0 = \hat{\theta}^\top \tilde{y}_i - \eta_i \hat{\theta}^\top C_y \hat{\theta} \quad \eta_i = \frac{\hat{\theta}^\top \tilde{y}_i}{\hat{\theta}^\top C_y \hat{\theta}}$$

From there

$$\tilde{y}_i - \hat{y}_i = \frac{\hat{\theta}^\top \tilde{y}_i C_y \hat{\theta}}{\hat{\theta}^\top C_y \hat{\theta}}.$$

**Generalized Total Least Squares (GTLS) estimate.**
The original objective function can be written

\[ J_{GTLS} = \sum_{i=1}^{n} \hat{\theta}^T \tilde{y}_i \hat{\theta} \frac{C_y C_y^{-1} C_y \tilde{y}_i \hat{\theta}}{(\hat{\theta}^T C_y \hat{\theta})^2} = \hat{\theta}^T (\sum_{i=1}^{n} \tilde{y}_i \tilde{y}_i^T) \hat{\theta} = \hat{\theta}^T \tilde{Y}^T \tilde{Y} \hat{\theta} \]

because

\[ \sum_{i=1}^{n} \tilde{y}_i \tilde{y}_i^T = [\tilde{y}_1 \ldots \tilde{y}_n] \left( \begin{array}{c} \tilde{y}_1^T \\ \vdots \\ \tilde{y}_n^T \end{array} \right) = \tilde{Y}^T \tilde{Y} \]

are two matrices \( p \times n \) and \( n \times p \) and recall \( \tilde{Y}^T = [\tilde{y}_1 \ldots \tilde{y}_n] \).

Now, we can do the main part, the parameter estimation. The gradient in \( \hat{\theta} \) of the objective function must be equal zero.

\[ \frac{\partial J_{GTLS}}{\partial \hat{\theta}} = 0 \Rightarrow 2 \tilde{Y}^T \tilde{Y} \hat{\theta} \begin{pmatrix} \hat{\theta} & C_y \hat{\theta} \end{pmatrix} - 2(\hat{\theta}^T \tilde{Y}^T \tilde{Y} \hat{\theta}) C_y \hat{\theta} = 0 \]

which can be written

\[ \tilde{Y}^T \tilde{Y} \hat{\theta} = \lambda C_y \hat{\theta}, \quad \lambda = \frac{\hat{\theta}^T \tilde{Y}^T \tilde{Y} \hat{\theta}}{\hat{\theta}^T C_y \hat{\theta}}. \]

The \( \lambda \) is a scalar function and we have to find the smallest eigenvalue in order to minimize the objective function. The matrices \( A = \tilde{Y}^T \tilde{Y} \) and \( C_y \) are \( p \times p \) and symmetric positive definite because it is a full rank problem. Therefore, we can also look for the smallest singular value which is the same in this case like the eigenvalue.

In MATLAB there are two subroutines which can be used. The routine –eig– which does generalized eigenvalue decomposition


or you can use generalized singular value decomposition, –gsvd–.


The ancillary constraint \( ||\hat{\theta}|| = 1 \) is very easy to satisfy after the solution is found. The intercept \( \hat{\alpha} \) is recovered as in the TLS. It can be shown that the covariance of
the parameters $\hat{\theta}$ is

$$C_{\hat{\theta}} = \hat{\sigma}^2 (\tilde{Y}^\top \tilde{Y} - \lambda_{\text{min}} C_y)^{-1}$$

where we kept in mind that the problem is full rank and substitute the estimates of the centered $y_i$-s. The variance of $\theta$ is defined as

$$\hat{\sigma}^2 = \frac{\lambda_{\text{min}}}{n - p} = \frac{\tilde{\sigma}_p^2}{n - p}.$$ 

The TLS and the generalized TLS were solved in different ways, but in the same condition (that is, for TLS) will give the same result. Applying GTLS to TLS gives

$$\tilde{Y}^\top \tilde{Y} \hat{\theta} = \lambda \hat{\theta}$$

while the other, TLS solution is

$$\tilde{Y} \hat{\theta} = 0.$$

The solution, which also finds the estimated input variables $\hat{y}_i$ for the $n$ points, is

1. the eigenvector corresponding to the smallest eigenvalue $\lambda_{\text{min}}$ of $\tilde{Y}^\top \tilde{Y}$; or
2. the projection into the null space of $\tilde{Y}$ of $\hat{\theta}$ which is the vector $v_p$.

As we have seen in the linear algebra lecture, this eigenvector is the vector $v_p$.

The ordinary LS can be transformed to GTLS with a rank deficient covariance matrix, where in our case only the $z$ has a variance. While this is not recommended, the GTLS will give the OLS solution.