Given the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^m \times \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{b} \in \mathbb{R}^m$, find the array $\boldsymbol{x}$ which verifies the first c equations and best-fits the following $m - c$ (i.e., in the sense of least-square).

The following partitioning for the matrices/arrays can be defined

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix} \quad \text{with} \quad \boldsymbol{A}_1 \in \mathbb{R}^c \times \mathbb{R}^n \quad \text{and} \quad \boldsymbol{A}_2 \in \mathbb{R}^{m-c} \times \mathbb{R}^n,$$

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} \quad \text{with} \quad \boldsymbol{b}_1 \in \mathbb{R}^c \quad \text{and} \quad \boldsymbol{b}_2 \in \mathbb{R}^{m-c}.$$

(The array $\boldsymbol{x}$ need not be partitioned as well.)

Defined the residual as $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$, the task consists in minimizing the quantity $\frac{1}{2}\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r}$, provided $\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{b}_1$ holds. In other words, a *constrained optimization problem* must be solved, where the constrain is the quation

$$\boldsymbol{A}_1\boldsymbol{x} - \boldsymbol{b}_1 = 0, \tag{1}$$

and the quantity to be minimezed is $\frac{1}{2}\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r}$.

The Lagrange function is

$$\begin{aligned} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) &= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}_1\boldsymbol{x} - \boldsymbol{b}_1) \\ &= \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \frac{1}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{\lambda}, \end{aligned}$$

and its partial derivatives with respect to $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ are

$$\frac{\partial\Phi(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial\boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{\lambda} \tag{2}$$

$$\frac{\partial\Phi(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}^{\mathrm{T}}} = \boldsymbol{A}_1\boldsymbol{x} - \boldsymbol{b}_1$$

Setting eq. (2) to zero, and solving with respect to $\boldsymbol{x}$, yields

$$\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{\lambda}). \tag{3}$$

Inserting eq. (3) into eq. (1), and rearranging the result, leads to

$$\boldsymbol{A}_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{b}_1 - \boldsymbol{A}_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b},$$

which can be solved for $\boldsymbol{\lambda}$, whose value can be substituted back into eq. (3), eventually obtaining the solution,

$$\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{A}_1^{\mathrm{T}}\underbrace{(\boldsymbol{A}_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}_1^{\mathrm{T}})^{-1}(\boldsymbol{b}_1 - \boldsymbol{A}_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b})}_{\boldsymbol{\lambda}}).$$