4: SLEs with Rectangular Matrix

Gaussian Elimination

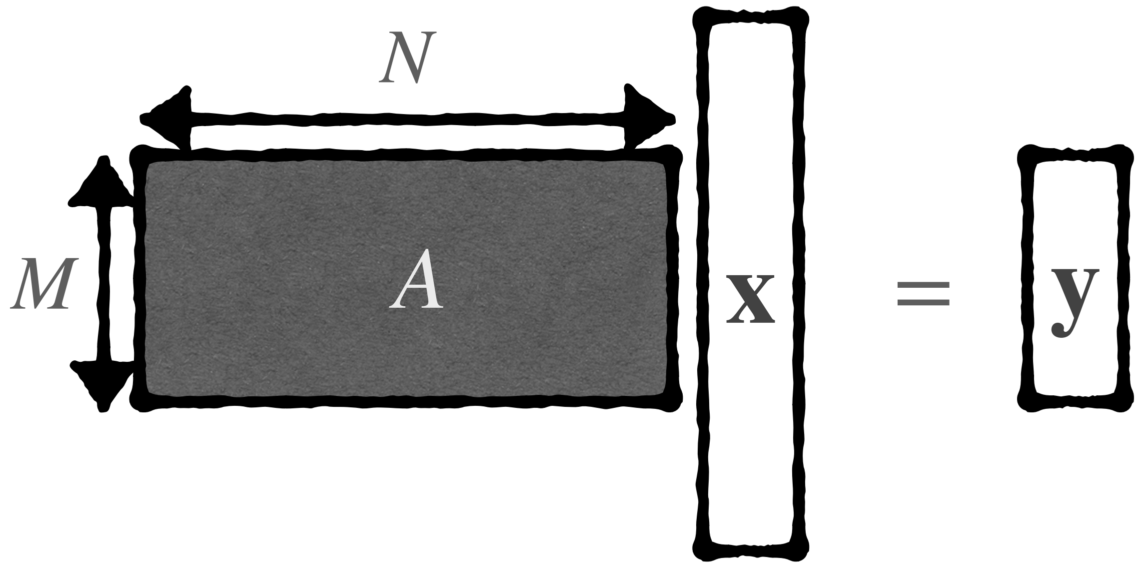
Matrix Kernel

Mikhail Romanov

Rectangular SLEs

Vars > # Data
Kernel

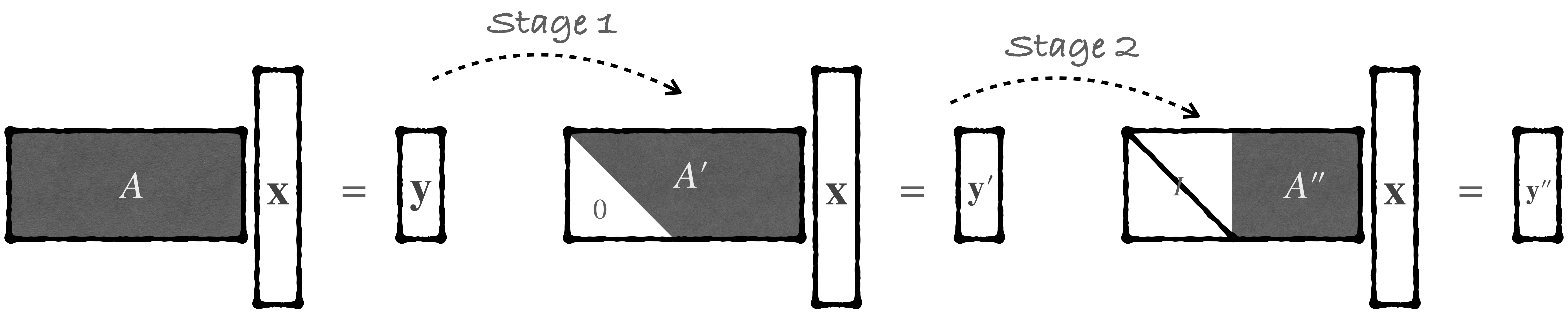
SLE: More variables than data



In this case $N > M$.

* More variables than data
* More freedom degrees than limitations
* There will be unfixed freedom degrees in the solution
* The solution will be non-unique

SLE: More variables than data



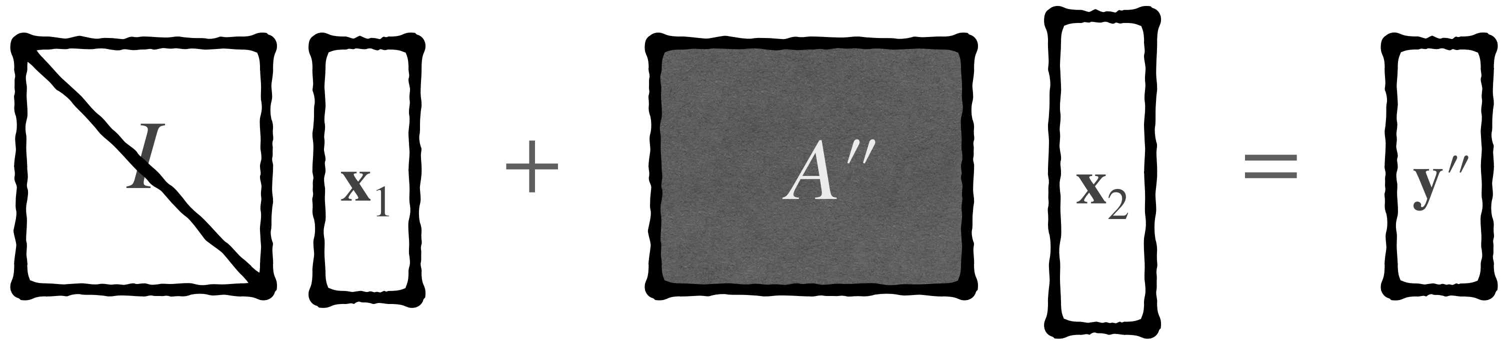
Let's try to use the same Gauss Algorithm

This is what happens if everything goes fine

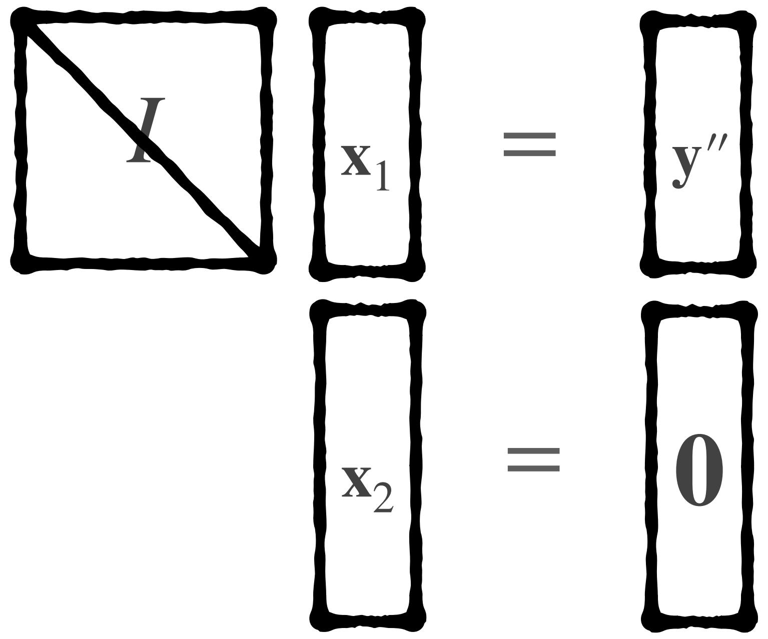
(No zeros on diagonal)

This is what happens if everything goes fine

(No zeros on diagonal)



One solution



Thus, we have found one solution $\mathbf{x}_p$

Theorem: All the solutions $A\mathbf{x} = \mathbf{y}$ are given by $\mathbf{x}_0 + \mathbf{x}_p$, where $\mathbf{x}_0$ are all solutions of $A\mathbf{x}_0 = \mathbf{0}$, and $\mathbf{x}_p$ is one solution if $A\mathbf{x}_p = \mathbf{y}$.

Now consider null-space of matrix A:

$$A\mathbf{x}_0 = \mathbf{0}$$

Then every vector $\mathbf{x}_p + \mathbf{x}_0$ is solution of $A\mathbf{x} = \mathbf{y}$:

$$A(\mathbf{x}_p + \mathbf{x}_0) = A\mathbf{x}_p + \underbrace{A\mathbf{x}_0}_{\mathbf{0}} = \mathbf{y}$$

The question is: are all of the solutions covered by $\mathbf{x}_p + \mathbf{x}_0$?

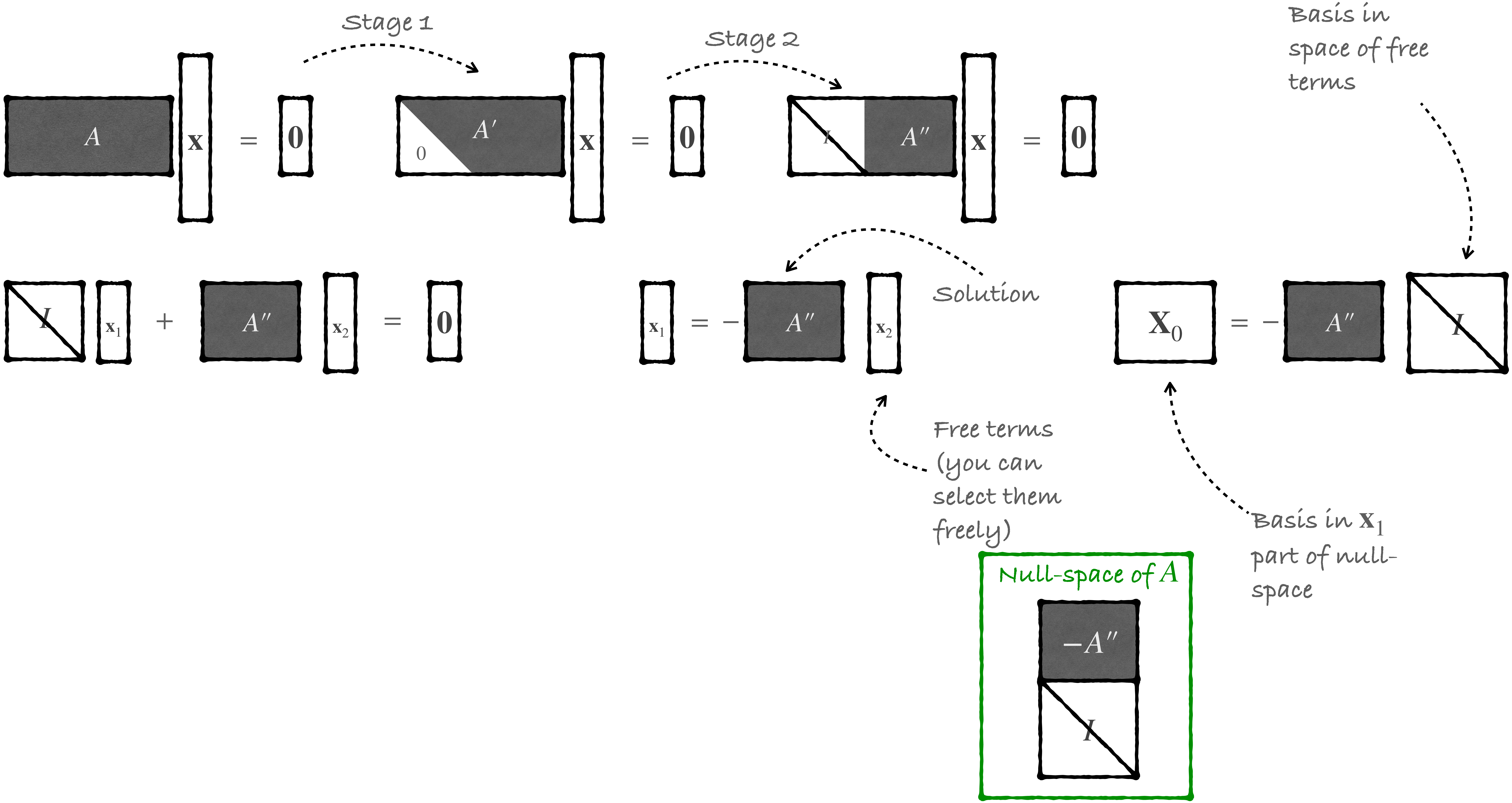
Consider some solution $\mathbf{x}_2$.

$$\underbrace{A\mathbf{x}_p}_{\mathbf{y}} - \underbrace{A\mathbf{x}_2}_{\mathbf{y}} = \mathbf{0}.$$

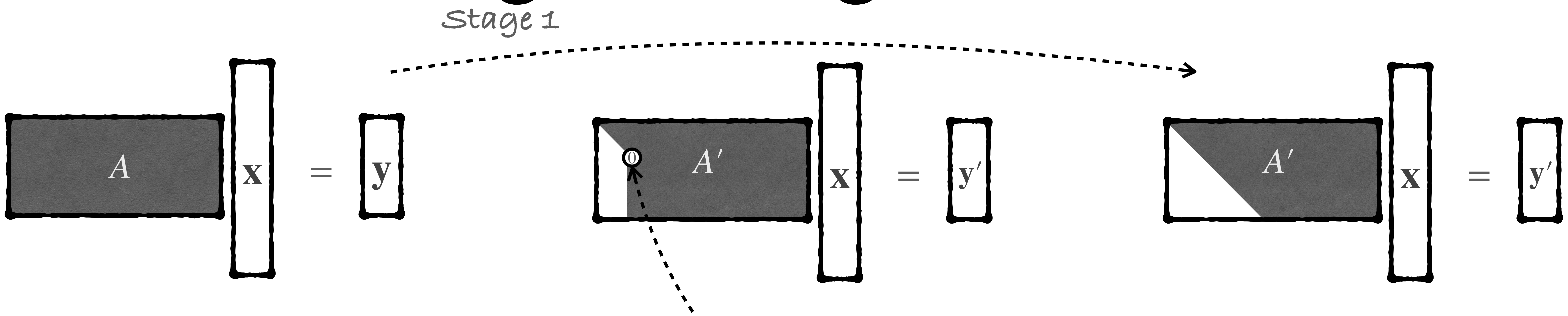
Hence, $A(\mathbf{x}_p - \mathbf{x}_2) = \mathbf{0}$.

Thus, to find all the solutions, one has to find one partial solution and null-space of A.

Kernel of matrix



What could go wrong?



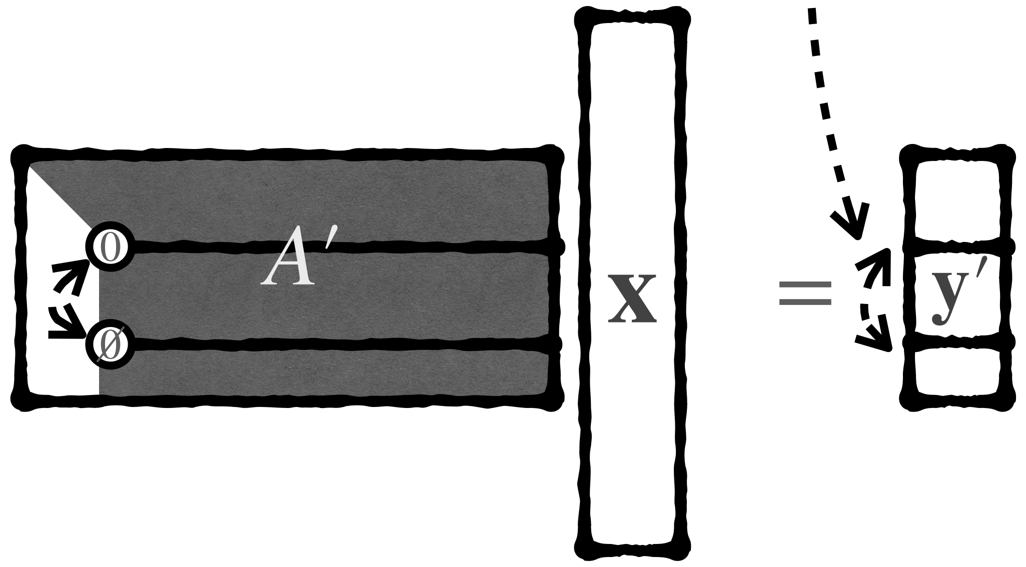
Emergencies may occur only at stage #1

Throughout the process we encountered zero on the main diagonal

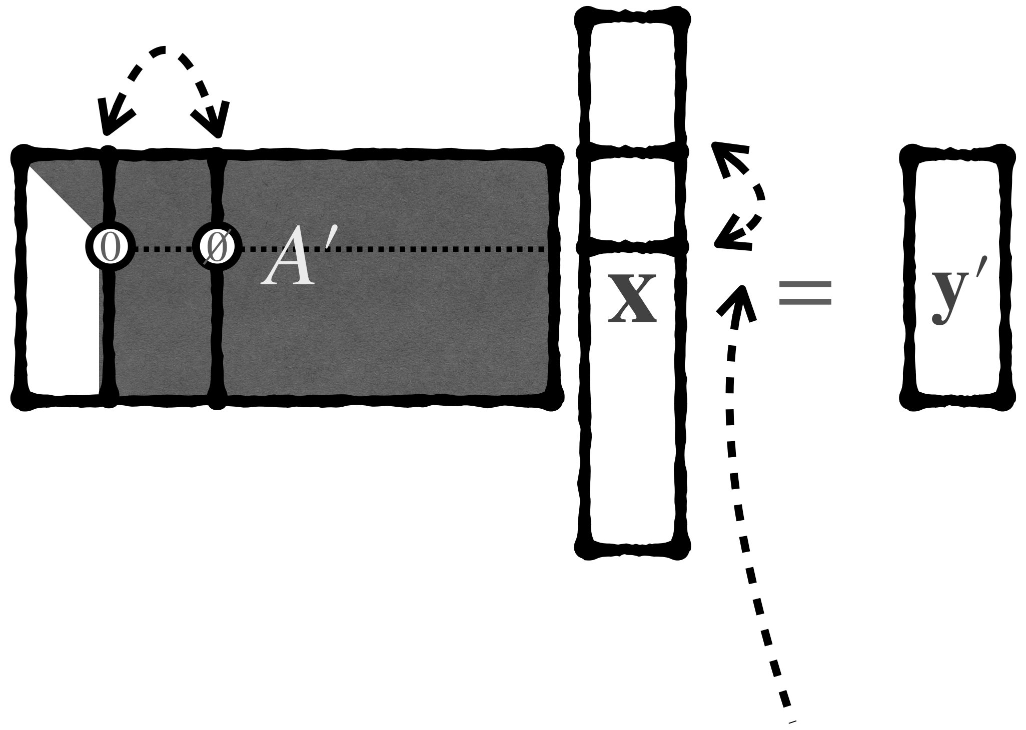
Note that coordinates in $\mathbf{y}$ swap then!

How to resolve:

Step 1: try to swap equations with the one that has non-zero in that place.



Step 2: try to swap the variables with the one that has non-zero there



No success?

- Left part is strictly 0
- If corresponding $y = 0$, this equation Is duplicate
- Otherwise, the system is inconsistent (no solution)

Note that coordinates in $\mathbf{x}$ swap then!

Rectangular SLEs

Vars < # Data
Least Squares Solution

SLE: More data than variables

Most likely INCONSISTENT

In case you are lucky — some equations are duplicate

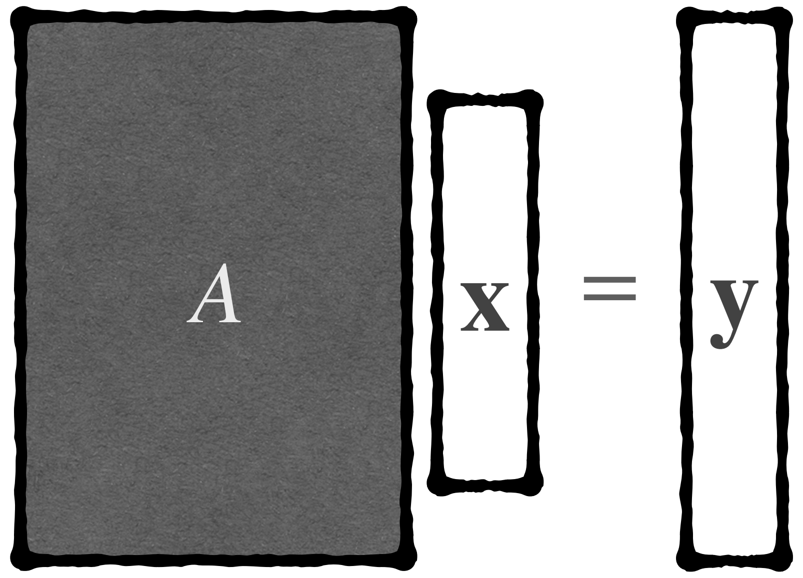
Otherwise you cannot solve it.

View SLE as degrees of freedom $\mathbf{x}$ fitting the data $\mathbf{y}$.

To fix one datum from $\mathbf{y}$ — remove one freedom.

Extra data cannot be fitted

But we can approximate it.



$$A\mathbf{x} = \mathbf{y}$$

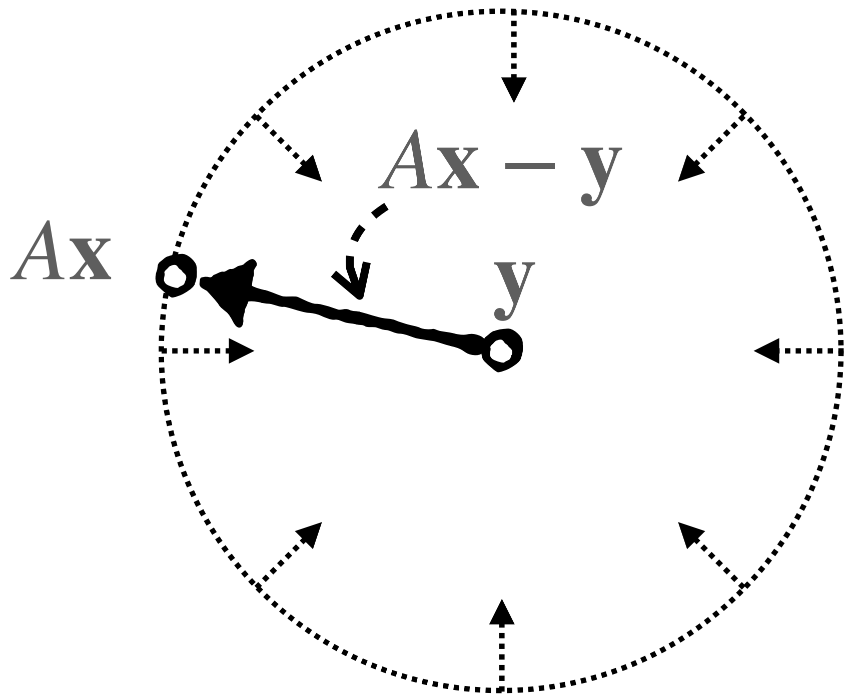
$$A\mathbf{x} - \mathbf{y} = \mathbf{0}$$

We cannot solve it exactly

But we can minimise the misfit:

$$\underbrace{\|A\mathbf{x} - \mathbf{y}\|} \to 0$$

Distance between $A\mathbf{x}$ and $\mathbf{y}$.



The closer $A\mathbf{x}$ to $\mathbf{y}$ is, the better is the guess $\mathbf{x}$

$$\|A\mathbf{x} - \mathbf{y}\| \to \min$$

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T \mathbf{y}$$

Memoization: Not derivation!

$$A\mathbf{x} = \mathbf{y}$$

$$A^T A\mathbf{x} = A^T \mathbf{y}$$

Square matrix

Matrix Derivatives

About the derivatives

Def: $$f' = \frac{\partial f}{\partial x} = \lim_{dx \to 0} \frac{f(x + dx) - f(x)}{dx}$$

Rules: $$(f(x) + g(x))' = f'(x) + g'(x)$$

$$(f(x) + g(x))' = f'(x) + g'(x)$$

Matrix Derivative

If you have a minimisation problem:

$$\operatorname*{argmin}_{\mathbf{x}} f(\mathbf{x})$$

A way to solve it is to find extremums,

The points, where the function does not change:

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}$$

All the coordinates

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

This thing is called <u>Gradient</u>

We need to set it to zero

Example: $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A \mathbf{x}$

We can do this due to:

$$\frac{\partial f + g}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x}$$

$$\frac{\partial}{\partial x_i} \sum_{k=1}^{K} \sum_{l=1}^{K} x_k a_{kl} x_l = \sum_{k=1}^{K} \sum_{l=1}^{K} \frac{\partial}{\partial x_i} x_k a_{kl} x_l =$$

Derivative of product:

$$\frac{\partial f \cdot g}{\partial x} = g \cdot \frac{\partial f}{\partial x} + f \cdot \frac{\partial g}{\partial x}$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\frac{\partial}{\partial x_i} x_k \right) a_{kl} x_l + x_k a_{kl} \left(\frac{\partial}{\partial x_i} x_l \right) =$$

$$\frac{\partial x_k}{\partial x_i} = \delta_{ki} = \begin{cases} 0 & \text{if} \quad k \neq i \\ 1 & \text{if} \quad k = i \end{cases}$$

Derivative is 1 only if $k = i$, otherwise it is 0

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \delta_{ik} a_{kl} x_l + x_k a_{kl} \delta_{il} =$$

Just rewriting the expression above

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \delta_{ik} a_{kl} x_l + \sum_{k=1}^{K} \sum_{l=1}^{K} x_k a_{kl} \delta_{il} =$$

Just expanding the sum

In the sum throughout k only when $k = i$ we have a non-zero term

$$= \sum_{l=1}^{K} a_{il} x_l + \sum_{k=1}^{K} a_{ki} x_k = A\mathbf{x} + A^T\mathbf{x}$$

Least Squares Solution: Derivation

$$\|A\mathbf{x} - \mathbf{y}\|^2 \to 0$$

$$(A\mathbf{x} - \mathbf{y})^T(A\mathbf{x} - \mathbf{y}) \to 0$$

$$(A\mathbf{x} - \mathbf{y})^T(A\mathbf{x} - \mathbf{y}) \to \min_{\mathbf{x}}$$

$$\mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} - \mathbf{x}^T A^T \mathbf{b} + \mathbf{b}^T\mathbf{b} \to \min_{\mathbf{x}}$$

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} - \mathbf{x}^T A^T \mathbf{b} + \mathbf{b}^T\mathbf{b}\right) = 0$$

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T A^T A\mathbf{x} - \frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T A\mathbf{x} - \frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T A^T \mathbf{b} + \frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T\mathbf{b} = 0$$

$$\boxed{\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T A^T A\mathbf{x}}$$

$$\frac{\partial}{\partial x_i}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M} x_k a_{kl} a_{lm} x_m = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M} a_{kl} a_{lm}(\delta_{ki} x_m + x_k \delta_{im}) =$$

$$= \sum_{l=1}^{L}\sum_{m=1}^{M} a_{il} a_{lm} x_m + \sum_{k=1}^{K}\sum_{l=1}^{L} a_{kl} a_{li} x_k = A^T A\mathbf{x} + \left(\mathbf{x}^T A^T A\right)^T = 2A^T A\mathbf{x}$$

$$\boxed{\frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T A\mathbf{x} = \frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T A^T \mathbf{b}}$$

$$\frac{\partial}{\partial x_i}\mathbf{b}^T A\mathbf{x} = \frac{\partial}{\partial x_i}\sum_{k=1}^{K}\sum_{l=1}^{L} b_k a_{kl} x_l = \sum_{k=1}^{K} b_k a_{kl} \delta_{li} = A^T\mathbf{b}$$

$$\boxed{\frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T\mathbf{b} = \mathbf{0}}$$

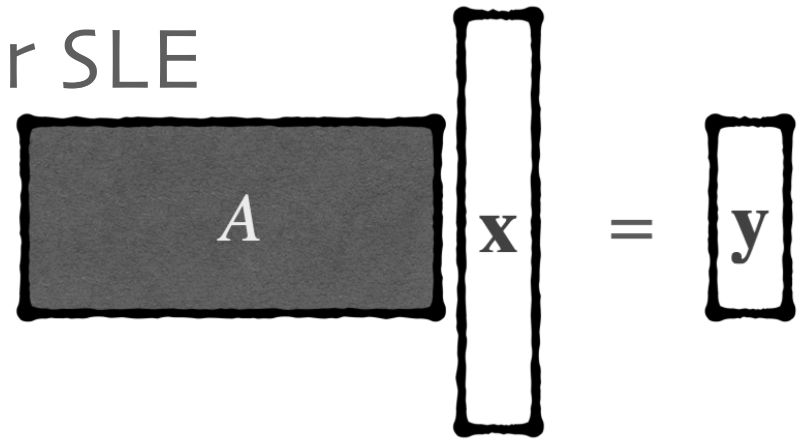
$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

$$2A^T A\mathbf{x} - 2A^T\mathbf{b} = \mathbf{0} \qquad A^T A\mathbf{x} = A^T\mathbf{b} \qquad \boxed{\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b}}$$

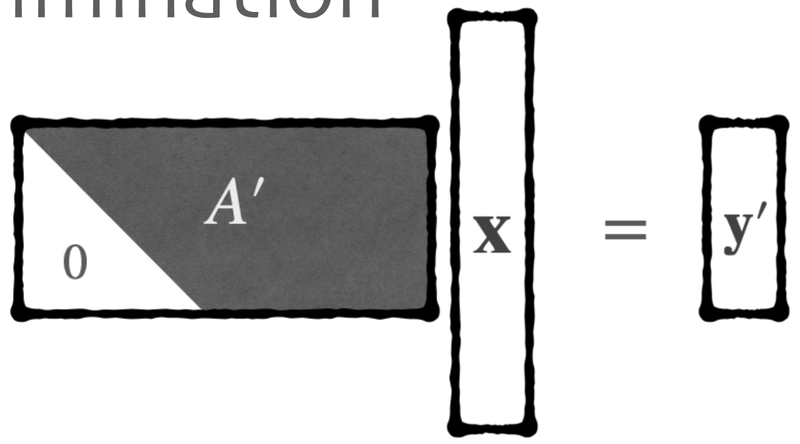
Least Squares Solution

Takeaways

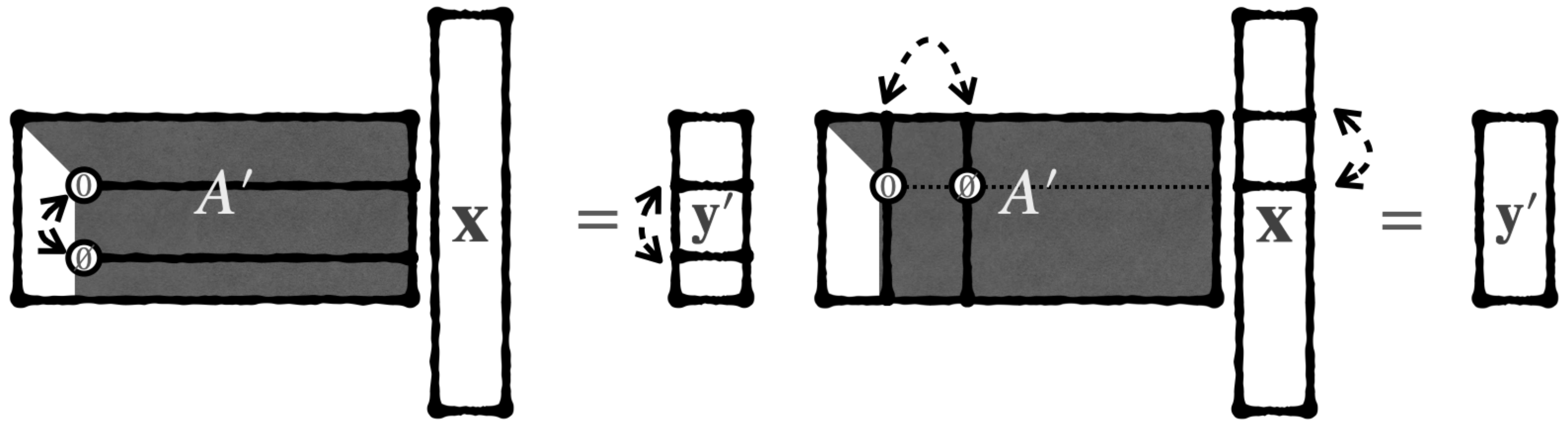
- Rectangular SLE



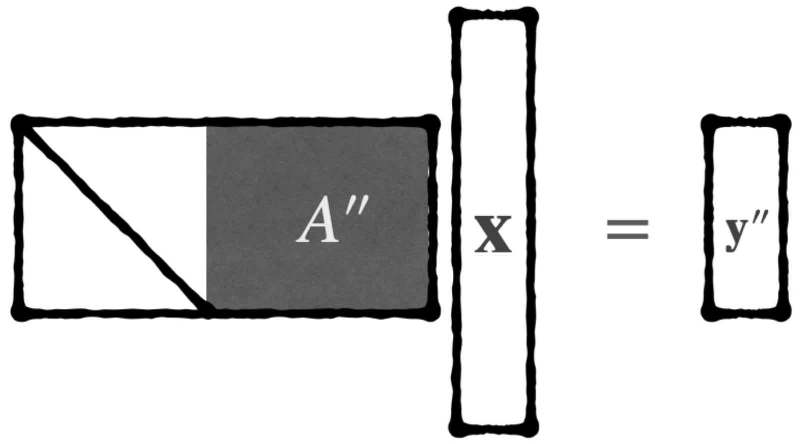
- Gaussian Elimination
  - Stage I



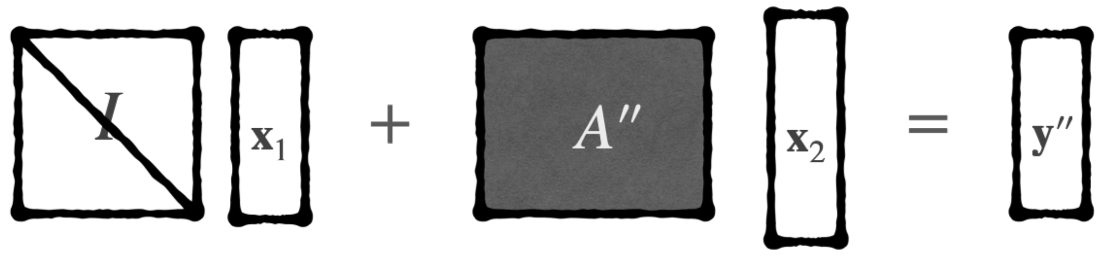
  - Issue resolving



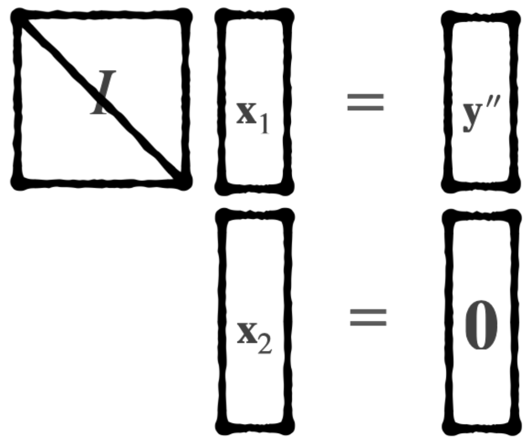
  - Stage II



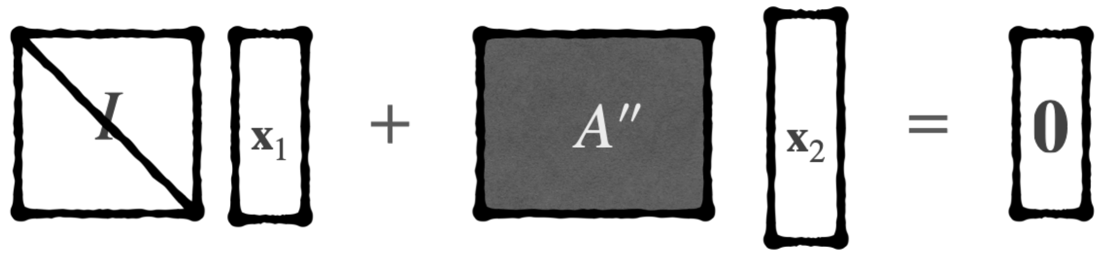
- Partial solution

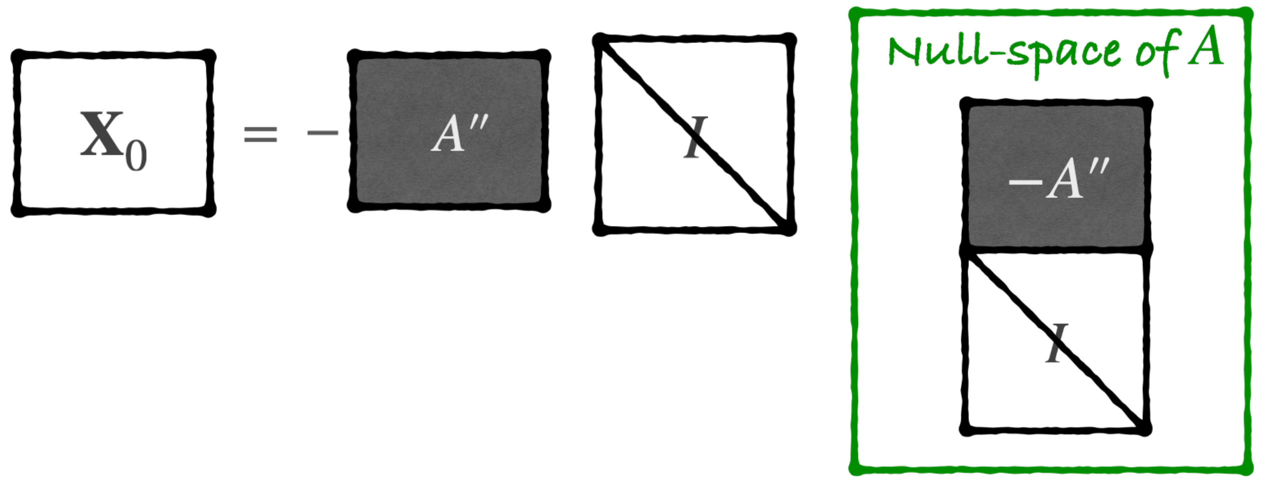


One solution



- Kernel $\left\{ \forall \mathbf{x_0} : A\mathbf{x_0} = \mathbf{0} \right\}$

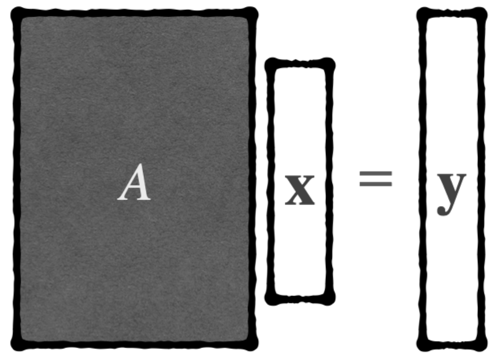




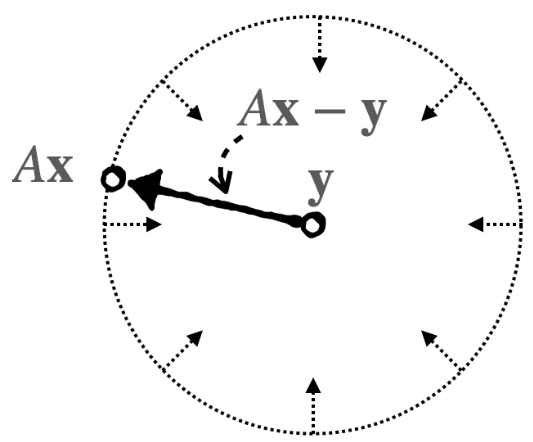
- General Solution:

$$A(\mathbf{x}_p + \mathbf{x}_0) = A\mathbf{x}_p + \underbrace{A\mathbf{x}_0}_{\mathbf{0}} = \mathbf{y}$$

- More data than variables



- No solutions (most likely)
- Closest solution formulation



$$\|A\mathbf{x} - \mathbf{y}\| \rightarrow \min$$

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|A\mathbf{x} - \mathbf{y}\| = \underset{\mathbf{x}}{\operatorname{argmin}} \|A\mathbf{x} - \mathbf{y}\|^2$$

- Closest solution derivation

$$\frac{\partial \|A\mathbf{x} - \mathbf{y}\|^2}{\partial \mathbf{x}} = \frac{\partial (A\mathbf{x} - \mathbf{y})^T (A\mathbf{x} - \mathbf{y})}{\partial \mathbf{x}} = \mathbf{0}$$

- Closest solution formula

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T \mathbf{y}$$