

# 1 Solving parameters using Gauss-Jordan elimination

Standard approach to solving determined sets of equations requires just enough constraints to ensure uniqueness of the solution, without it being overdetermined (which can lead to contradictions). In our case, the set of available information is greater than required number of equations that satisfy above conditions. Out of  $2^n$  possible binary vectors representing joint probability of events, we are forced to pick  $n$  that introduce the fewest error in further calculations. This approach can yield good results when done properly, but the question of which equations to choose remains unanswered. Preserving linear independence of vectors invoke additional complexity to the rules by which we decide the final set of equations.

Using Gauss-Jordan elimination, we can avoid this problem entirely, since it allows us to work with both overdetermined and underdetermined systems of equations. The order of equations is also taken into account, so in cases of contradictions, certain combinations are preferred to others. General rules of Gauss elimination apply, but are slightly modified to fit our product-equations (rather than linear equations). The outcome of our procedure for Gauss-Jordan elimination is a reduced row echelon form.

## 1.1 Example of Gauss-Jordan elimination

Let us work with this simple example written in a standard matrix form  $A \cdot X = b$ .

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad (1)$$

We will use symbols for vector  $b$ , so we can keep track of operations with constants. Turning that into augmented matrix  $[A|b]$  yields

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \quad (2)$$

Notice that this equation set is both contradictory ( $x_3$  can be calculated in two ways), and under-determined (There is no way to distinguish between  $x_1$  and  $x_2$  parameters). We will now perform Gauss-Jordan elimination steps in order to show that certain properties we care about (such as preserving preference of equations determined by their order) apply.

Since we do not focus on any particular column, so we will try to solve the whole equation. We will distinguish pivot elements with colors red (currently selected pivot element) and blue (previous pivot elements).

1. We choose the our first pivot element (in red), and use it to zero-out remaining coefficients in first column.

$$\left[ \begin{array}{cccc|c} \textcolor{red}{1} & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] r_2 = r_2 - r_1 \sim \left[ \begin{array}{cccc|c} \textcolor{blue}{1} & 1 & 1 & 0 & b_1 \\ 0 & 0 & -1 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \quad (3)$$

2. We select the second pivot element - note that no two pivot elements can share the same row. The first non-zero element that satisfies this condition is coefficient of  $x_3$  in second row. Choosing the first element from the top guarantees that the order of preference of equations is taken into account.

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 0 & 0 & -1 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] r_2 = r_2 \cdot (-1) \sim \left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \quad (4)$$

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \begin{array}{l} r_1 = r_1 - r_2 \\ r_4 = r_4 - r_2 \end{array} \sim \left[ \begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 0 & 0 & b_4 - (b_1 - b_2) \end{array} \right] \quad (5)$$

3. The last pivot element is going to be coefficient at  $x_4$  in fourth row. Since the remaining coefficients are all zeros in fourth column, no changes are made. We can simplify the values in new vector  $b$ . Notice that fourth row is a zero-vector - we can eliminate that from the equation set.

$$\left[ \begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 0 & 0 & b_4 - (b_1 - b_2) \end{array} \right] \sim \left[ \begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{array} \right] \quad (6)$$

Let's compare our end-result with the initial matrix:

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \quad (7)$$

As we can see,  $x_3$  was calculated using first and second row ( $b_1 - b_2$ ). Let's see what happens after we move third row on the top position, indicating that our preferred ordering of equation changed. (We expect now to calculate  $x_3$  solely by first row).

1. We choose the our first pivot element, and use it to zero-out remaining coefficients in first column.

$$\left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 1 & 0 & b_2 \\ 1 & 1 & 0 & 0 & b_3 \\ 0 & 0 & 0 & 1 & b_4 \end{array} \right] r_3 = r_3 - r_2 \sim \left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 1 & 0 & b_2 \\ 0 & 0 & -1 & 0 & b_3 - b_2 \\ 0 & 0 & 0 & 1 & b_1 \end{array} \right] \quad (8)$$

2. Again, no candidate for pivot element in second column, coefficient at  $x_3$  in first row is the next pivot element

$$\left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 1 & 0 & b_2 \\ 0 & 0 & -1 & 0 & b_3 - b_2 \\ 0 & 0 & 0 & 1 & b_1 \end{array} \right] \begin{array}{l} r_2 = r_2 - r_1 \\ r_3 = r_3 + r_1 \end{array} \sim \left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array} \right] \quad (9)$$

3. Last item fit for a pivot element is a coefficient at  $x_4$  in fourth row. After getting rid of zero vectors, we achieve the following reduced row echelon form matrix

$$\left[ \begin{array}{cccc|c} 0 & 0 & \textcolor{blue}{1} & 0 & b_1 \\ \textcolor{blue}{1} & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 + b_1 \\ 0 & 0 & 0 & \textcolor{red}{1} & b_4 \end{array} \right] \sim \left[ \begin{array}{cccc|c} 0 & 0 & \textcolor{blue}{1} & 0 & b_1 \\ \textcolor{blue}{1} & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & \textcolor{blue}{1} & b_4 \end{array} \right] \quad (10)$$

As expected,  $x_3$  was calculated using the most preferred set of equations, as dictated by their order. What this method does not take into account is the relative weight (or preference) of each row. As of yet, all we could rely on was simple ordering of equations, without using the information about quantity or frequency of each type of equation in our learning set. If our method was to provide that feature to us as well, we could talk about very complete and solid solution that can be expected to perform optimally.

## 1.2 Properties of zero vectors

In this section we will describe how solving equation set without any particular ordering does not DEPRIVE[FIXME] us from reproducing other ways to calculate given parameter. As we saw previously, different order of equation leads to different outcomes for certain parameters. This variety came from contradictions in equation set. Zero vectors that emerge during Gauss-Jordan elimination contain information about other ways to calculate given value. Instead of removing them in the process, we can store them, and utilize them later. Let's see how previous example holds to that theory.

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \end{array} \right] \sim \left[ \begin{array}{cccc|c} 1 & 1 & 0 & 0 & b_1 - (b_1 - b_2) \\ 0 & 0 & 1 & 0 & b_1 - b_2 \\ 0 & 0 & 0 & 1 & b_3 \\ \textcolor{blue}{0} & \textcolor{blue}{0} & \textcolor{blue}{0} & \textcolor{blue}{0} & b_4 - (b_1 - b_2) \end{array} \right] \sim \left[ \begin{array}{cccc|c} \textcolor{blue}{1} & 1 & 0 & 0 & b_2 \\ \textcolor{red}{0} & \textcolor{red}{0} & \textcolor{red}{1} & \textcolor{red}{0} & \textcolor{red}{b_1 - b_2} \\ 0 & 0 & 0 & 1 & b_3 \end{array} \right] \quad (11)$$

This particular order of equation lead to  $x_3$  being calculated from two top-most equations in a set. If we add our final solution for  $x_3$  (vector in red), to the zero vector we ought to remove in a penultimate step of our algorithm (vector in blue), we obtain previously abandoned solution:

$$[0 \ 0 \ 1 \ 0 \mid b_1 - b_2] + [0 \ 0 \ 0 \ 0 \mid b_4 - (b_1 - b_2)] = [0 \ 0 \ 1 \ 0 \mid b_4] \quad (12)$$

Using the same method we can start from the solution obtained after rearranging the order of equations in initial matrix.

$$\left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 1 & 0 & b_2 \\ 1 & 1 & 0 & 0 & b_3 \\ 0 & 0 & 0 & 1 & b_4 \end{array} \right] \sim \left[ \begin{array}{cccc|c} 0 & 0 & 1 & 0 & b_1 \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ \textcolor{blue}{0} & \textcolor{blue}{0} & \textcolor{blue}{0} & \textcolor{blue}{0} & b_3 - b_2 + b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array} \right] \sim \left[ \begin{array}{cccc|c} \textcolor{red}{0} & \textcolor{red}{0} & \textcolor{red}{1} & \textcolor{red}{0} & \textcolor{red}{b_1} \\ 1 & 1 & 0 & 0 & b_2 - b_1 \\ 0 & 0 & 0 & 1 & b_4 \end{array} \right] \quad (13)$$

Linear combination of two vectors again yields a different solution

$$[0 \ 0 \ 1 \ 0 \mid b_1] + (-1) \cdot [0 \ 0 \ 0 \ 0 \mid b_3 - b_2 + b_1] = [0 \ 0 \ 1 \ 0 \mid b_2 - b_3] \quad (14)$$

Above property is crucial in showing that although the method itself does not take relative frequency, or weight between the equations in a set, it does provide us with a collection of solutions for given parameter. This way we can iterate over the set of possible solutions and choose the one that minimizes the relative error, or take a weighted average as our solution.

Let's treat fourth column that responds to quantity of each equation in a dataset. Whole equation set adds up to 2500 records. We can calculate a frequency for each of the probabilities, and treat them as weights.

	probability	quantity	frequency $Fq(b_i)$
0 0 1 0	$b_1$	980	0.392
1 1 1 0	$b_2$	760	0.304
1 1 0 0	$b_3$	440	0.176
0 0 0 1	$b_4$	320	0.128

(15)

Using this data we can come up with few heuristics for calculating final value of  $x_3$  or compare different solutions. Let's propose a fitness function for a solution:

$$F(s) = \frac{\prod_{b_i \in s} Fq(b_i)}{\sum_{b_i \in s} 1}, \quad (16)$$

, where  $b_i \in s$  is true when  $b_i$  is taken into account (adding or subtracting) in given solution.

	probability	fitness function $F(s)$
$[0 \ 0 \ 1 \ 0]_1$	$b_1$	0.392
$[0 \ 0 \ 1 \ 0]_2$	$b_2 - b_3$	$\frac{0.304 \cdot 0.176}{2} = 0.24$

(17)

At this point we can pick a solution with a higher fitness value, or take weighted average of each solution as our final answer:

	probability
$[0 \ 0 \ 1 \ 0]$	$\frac{0.392 \cdot b_1 + 0.24 \cdot (b_2 - b_3)}{0.392 + 0.24}$

(18)

Let's look at the example of an equation set with multiple zero vectors:

$$\left[ \begin{array}{cccc|c} 1 & 1 & 0 & 1 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & 0 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \\ 1 & 0 & 0 & 0 & b_5 \\ 1 & 0 & 0 & 1 & b_6 \end{array} \right] \quad (19)$$

This equation set is of course overconstrained. We can expect to obtain at least two zero vectors

after the Gauss-Jordan elimination steps.

$$\left[ \begin{array}{cccc|c} 1 & 1 & 0 & 1 & b_1 \\ 1 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & 0 & b_3 \\ 0 & 0 & 1 & 0 & b_4 \\ 1 & 0 & 0 & 0 & b_5 \\ 1 & 0 & 0 & 1 & b_6 \end{array} \right] \sim \left[ \begin{array}{cccc|c} 1 & 0 & 0 & 0 & b_2 - b_3 + b_4 \\ 0 & 0 & 0 & 1 & b_1 - b_2 \\ 0 & 1 & 0 & 0 & b_3 - b_4 \\ 0 & 0 & 1 & 0 & b_4 \\ 0 & 0 & 0 & 0 & -b_2 + b_3 - b_4 + b_5 \\ 0 & 0 & 0 & 0 & -b_1 + b_3 - b_4 + b_6 \end{array} \right] \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{matrix} \quad (20)$$

Let's mark each vector in a reduced row echelon form ( $v_1 \dots v_6$ ). As we had shown previously, we can use zero vectors to obtain different solutions to parameters  $x_1 \dots x_4$ , for example:

	combination	value
$[1 \ 0 \ 0 \ 0]_1$	$v_1$	$b_2 - b_3 + b_4$
$[1 \ 0 \ 0 \ 0]_2$	$v_1 + v_5$	$b_5$
$[1 \ 0 \ 0 \ 0]_3$	$v_1 + v_6$	$-b_1 + b_2 + b_6$
$[0 \ 1 \ 0 \ 0]_1$	$v_3$	$b_3 - b_4$
$[0 \ 1 \ 0 \ 0]_2$	$v_3 - v_5$	$b_2 - b_5$
$[0 \ 1 \ 0 \ 0]_3$	$v_3 - v_6$	$b_1 - b_6$
...	...	...
$[0 \ 0 \ 0 \ 1]_1$	$v_2$	$b_1 - b_2$
$[0 \ 0 \ 0 \ 1]_2$	$v_2 + v_6 - v_5$	$b_6 - b_5$
...	...	...

(21)

Notice that second solution for  $x_4$  (in red above) requires two zero vectors to find an efficient solution.

We would like to propose a conjecture describing relationship of possible solutions with zero vectors in reduced row echelon form matrix.

### Zero-vector conjecture

Every possible solution for given parameter can be obtained as a linear combination of a solution vector from Gauss-Jordan elimination, and the zero vectors, that is

$$\forall_{s \in S} \exists_{a \in V} s = [1, a_1, a_2, \dots, a_n] \times [s_0, z_1, z_2, \dots, z_n] = s_0 + a_1 \cdot z_1 + a_2 \cdot z_2 + \dots + a_n \cdot z_n \quad (22)$$

where  $s \in S$  is a solution vector  $s$  for given parameter from a solution space  $S$

$a \in V$  is a vector of coefficients from a vector space  $V$ , over a field  $\mathbb{R}$

$s_0$  is a first solution (also a vector over a field  $\mathbb{R}$ ) for given parameter, as obtained from Gauss-Jordan elimination

and  $z_i$  is the  $i$ -th zero vector obtained from Gauss-Jordan elimination ( $i \in 1 \dots n$ ).

$n$  is the number of zero vectors in reduced row echelon form.

### Proof

If the equation set is determined, each parameter has a unique solution. Since in that case there are no zero vectors,  $n = 0 \Rightarrow s = [1] \times [s_0] = s_0$ .

Similar case would emerge when the equation set is strictly underdetermined (not every parameter has a unique solution, but no zero vectors appear in reduced row echelon form either).

Third case would be equation sets with over-constraints, which are of our interest here since they produce zero vectors after Gauss-Jordan elimination.

First, let's define two terms we will later use:

### Linear combination of equation set

Linear combination of equation set can be interpreted as a function

$$f : \mathbb{M}_{m \times n} \rightarrow \mathbb{M}_{m \times n} \quad (23)$$

where  $\mathbb{M}_{m \times n}$  is a space of matrices of size  $m \times n$ .

Additionally every such function  $f$  is equivalent to some matrix  $F$ , that is

$$f(A_0) = F \cdot A_0 \quad (24)$$

Example: Equation 20 could also be written as

$$\begin{bmatrix} 0 & 1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 & 1 & | & b_1 \\ 1 & 1 & 0 & 0 & | & b_2 \\ 0 & 1 & 1 & 0 & | & b_3 \\ 0 & 0 & 1 & 0 & | & b_4 \\ 1 & 0 & 0 & 0 & | & b_5 \\ 1 & 0 & 0 & 1 & | & b_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & | & b_2 - b_3 + b_4 \\ 0 & 0 & 0 & 1 & | & b_1 - b_2 \\ 0 & 1 & 0 & 0 & | & b_3 - b_4 \\ 0 & 0 & 1 & 0 & | & b_4 \\ 0 & 0 & 0 & 0 & | & -b_2 + b_3 - b_4 + b_5 \\ 0 & 0 & 0 & 0 & | & -b_1 + b_3 - b_4 + b_6 \end{bmatrix} \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{matrix} \quad (25)$$

in which case the leftmost matrix would be our linear combination of Gauss-Jordan elimination steps.

### Solution vector

Solution vector is a single row in a matrix (usually obtained by linear combination of equation set), directly solving given parameter  $k$  (vector  $[0 \dots 0 \ 1 \ 0 \dots 0]$  with 1 at the  $k$ -th place). Example: Row  $v_2$  in equation 25 unambiguously gives solution to parameter  $x_4$ , thus vector  $[0 \ 0 \ 0 \ 1 \ | \ b_1 - b_2]$  is the solution vector of  $x_4$ .

We will prove now the conjecture in question. Let  $\mathbf{A}$  be the original equation set, and  $\mathbf{B}$  - the equation set after Gauss-Jordan elimination (reduced row echelon form). Let's say that given parameter  $x_m$  can be calculated using two different linear combinations of vectors in the original equation. Let's call these  $L_0$  and  $L_k$ , where  $L_0$  is linear combination equivalent to Gauss-Jordan elimination ( $\mathbf{B} = L_0 \cdot \mathbf{A}$ ). Let's assume that  $s_0$  is the solution vector for parameter  $x_m$  as obtained from Gauss-Jordan elimination (that is  $s_0 \in \mathbf{B}$ )

Solution  $s_0$  is a vector in in  $\mathbf{B}$ .

Solution  $s_k$  is a vector in  $L_k \cdot \mathbf{A}$ , but it is not a vector  $\mathbf{B}$  (otherwise we would have linearly dependent non-zero vectors in  $\mathbf{B}$  which does not happen after Gauss-Jordan elimination steps).

We can show that  $s_k$  can also be obtained as linear combination of vectors from  $\mathbf{B}$  using only  $s_0$  and the zero vectors, by splitting the conjecture into two parts:

1. Vector  $s_k$  can be obtained as a linear combination of vectors from  $\mathbf{B}$

Since  $L_0$  is determined by Gauss-Jordan elimination, which in turn use only elementary row operations,  $L_0$  is an invertible matrix. Thus  $L_0^{-1}$  exists.

$$L_k \cdot L_0^{-1} \cdot \mathbf{B} = L_k \cdot L_0^{-1} \cdot L_0 \cdot \mathbf{A} = L_k \cdot \mathbf{A} \quad (26)$$

Because  $s_k$  is a vector in  $L_k \cdot \mathbf{A}$  then  $s_k$  is also a vector in  $L_k \cdot L_0^{-1} \cdot \mathbf{B}$ . In that case we can use a linear combination  $L_k \cdot L_0^{-1}$  to go from solution  $s_0$  to  $s_k$ .

2. Vector  $s_k$  in  $L_k \cdot L_0^{-1} \cdot \mathbf{A}$  is a linear combination of vectors no other than  $s_0$  and the zero vectors in  $\mathbf{B}$

(Quite foggy argument I'm not satisfied with yet)

Because  $\mathbf{B}$  is a reduced row echelon form, it holds the following property: no two non-zero coefficients in reduced row echelon form share the same column or a row. Let's assume that vector  $s_k$  is calculated using two non-zero vectors in it's linear combination from  $\mathbf{B}$  to  $L_k \cdot \mathbf{A}$ . In that case,  $s_0$  has to be appear in a linear combination with a non-zero coefficient since no other non-zero vector can produce a 1 in  $m$ -th column. Additionally, any linear combination involving any two non-zero vectors from  $\mathbf{B}$  with both coefficients other than 0 will produce a vector with at least two parameters other than 0. Such vector would not be a solution vector, which  $s_k$  is. Contradiction.

Finding such linear combination is no trivial task. The solution space is infinite, and virtually any linear combination of zero vectors can be added to any non-zero vector, giving us an valid solution (although the combination would be very inefficient in most cases)