Validity of natural isotope abundance correction for metabolic flux analysis

Roland Nilsson

April 30, 2020

Abstract
A pervasive issue in stable isotope tracing and metabolic flux analysis is the presence of naturally occurring isotopes such as \(^{13}\)C. For mass isotopomer distributions (MIDs) measured by mass spectrometry, it is common practice to correct for natural isotopes within metabolites of interest using a particular linear transform based on binomial distributions. However, the origin and mathematical derivation of this transform is rather obscure, and it may be difficult for nonexperts to understand precisely how to interpret the resulting corrected MIDs. Moreover, corrected MIDs are often used to fit metabolic network models and infer metabolic fluxes, which implicitly assumes that corrected MIDs will yield the same flux solution as the actual observed MIDs. Yet, there seems to be no published proof of this important property. Here, we provide a detailed derivation of the MID-correcting linear transform, reflecting its historical development, and describe some interesting properties. We also provide a proof that for metabolic flux analysis on noise-free MID data at steady state, observed and corrected MIDs indeed yield the same flux solution. On the other hand, for noisy MID data, the flux solution will generally differ between the two representations.

1 Introduction
When analyzing data from stable isotope tracing experiments, the presence of naturally occurring isotopes can be problematic. For example, carbon-13 (\(^{13}\)C) accounts for about 1% of all carbon in the biosphere [Berglund and Wieser, 2011], and when measuring metabolic products we must somehow correct for this natural \(^{13}\)C to determine the amount of \(^{13}\)C derived from a tracer [Buescher et al., 2015]. Obviously, there is no physical difference between naturally occurring and tracer-derived \(^{13}\)C, so this correction can only be made statistically. To make this notion precise, we will need some definitions. For mass spectrometry data, an \(n\)-carbon compound can be divided into \(n+1\) mass isotopomers (also called isotopologues), where mass isotopomer \(i\) represents the set of molecules of the compound that contain exactly \(i\) \(^{13}\)C atoms \((i = 0, 1, \ldots, n)\), regardless of their position in the molecule. We will consider only \(^{13}\)C in this paper, but other isotopes such as \(^{15}\)N or \(^2\)H can be handled similarly. The associated mass isotopomer distribution (MID) is an \((n+1)\)-vector \(x = (x_0 \ x_1 \ \cdots \ x_n)^\top\) where \(x_i\) is the fraction of molecules having \(i\) \(^{13}\)C atoms. (Throughout, we will index MID vectors and related matrices starting from zero for convenience.) Because natural \(^{13}\)C atoms occur randomly and independently with a probability \(p \approx 0.01\), the “natural” mass isotopomer distribution (MID) of an \(n\)-carbon compound is a vector \(\beta^n\) whose elements follow the binomial distribution

\[
\beta_i^n = \left(\begin{array}{c} n \\ i \end{array}\right) p^i (1 - p)^{n-i}, \quad i = 0, 1, \ldots, n.
\]
In isotope tracing experiments, natural-derived and tracer-derived $^{13}$C inevitably mix to form the observed MIDs. An MID correction method is then a function that maps any observed MID $x$ to a hypothetical MID $y$ that would have been observed if natural $^{13}$C did not exist. We refer to $y$ as a corrected MID. In particular, MID correction must map the natural MID $\beta^n$ to the unit vector $e^{n,0} = (1 \ 0 \ \cdots \ 0)^\top$. Over the years, several methods and software packages have been developed for MID correction [Millard et al., 2012, Jungreuthmayer et al., 2015, Su et al., 2017, Millard et al., 2019], providing various features such as support for multiple isotopes and compound fragmentation. They are all based on a particular class of linear transforms $T^n, n = 1, 2, \ldots$ that relate an $n$-carbon MID $x$ to the corrected MID $y$ as $x = T^n y$, where $T^n$ is the $(n + 1) \times (n + 1)$ matrix

$$T^n = \begin{pmatrix} \beta_0^n & 0 & \cdots & 0 \\ \beta_1^n & \beta_0^{n-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n^n & \beta_{n-1}^n & \cdots & \beta_0^n \end{pmatrix} \tag{1}$$

MID correction then simply amounts to solve this system for $y = (T^n)^{-1} x$ (the matrix $T^n$ is always invertible). However, it is difficult to find literature explaining how this particular transform is derived and demonstrating why the MID $y$ is indeed a corrected MID, as defined above. The transform $T^n$ was developed in a series of technical studies going back as far the 1960’s [Brauman, 1966], but these may be inaccessible to non-experts, and so the origin and underlying logic of the transform $T^n$ remains somewhat obscure. An excellent survey of the historical development of MID correction was given by Midani et al. [2017], but did not include a mathematical derivation of the transform $T^n$. In section 2 of this paper, we provide such a derivation, reflecting the historical development. We then discuss some interesting properties of $T^n$ in section 3.

Another important question arises when the corrected MIDs $y$ are used for metabolic flux analysis (MFA). Performing MFA on corrected MIDs is attractive since we need not keep track of natural isotopes, simplifying the analysis [Buescher et al., 2015]. But, cellular metabolism obviously acts on the ”uncorrected” MIDs $x$ that physically occur, and therefore MFA on these uncorrected MIDs (where ”unlabeled” molecules have binomial MIDs $\beta^n$) must be regarded as the ground truth. Therefore, to ensure that MFA on the corrected MIDs $y$ is valid, we must first establish that this always yields the same results as MFA on the actual MIDs $x$. To our knowledge there is no published proof of this proposition. We provide such a proof for the case of noise-free data at isotopic steady-state in section 4, and conclude by discussing some remaining open questions.

## 2 Derivation of the MID correction

To understand the origin of the MID correcting transform $T^n$, we here provide a derivation reflecting its historical development [Midani et al., 2017]. We will first review mixture models of MIDs, which constitute a related but somewhat different type of ”correction” that yields the fractional abundance of one or more tracers in a metabolite pool, thereby accounting for natural $^{13}$C. We then describe a general model for the MID of labeled compounds that occur in mixture models, and finally show how these models can be generalized to obtain the transform $T^n$. 


2.1 Linear mixtures of tracers

We first consider the simple case where an \( n \)-carbon tracer with known MID \( x^1 \) is mixed with an endogenous source of the same compound, which has natural MID \( x^0 = \beta^0 \) (Figure 1). At steady state, the observed (measured) MID \( x \) is then a linear mixture

\[
x = a_0 x^0 + a_1 x^1
\]

where \( a_0, a_1 \) are the mixture fractions (satisfying \( a_0 + a_1 = 1 \)). This situation typically arises in isotope tracing in vivo, where \( x^0 \) is known as the tracee [Cobelli et al., 1992]. Provided that we know the tracer MID \( x^1 \) and can measure the MID \( x \), the above yields \( n + 1 \) equations and two unknowns, and we can solve for the mixture coefficients \( a = (a_0, a_1) \) using a least-squares method [Brauman, 1966]. For the example data given in Figure 1, this gives \( a = (0.85, 0.15) \), indicating that the tracer constitutes 15% of the observed mixture, after "correction" for natural \(^{13}\)C. As expected, \( a_1 < x_1 = 0.191 \), since natural \(^{13}\)C also contributes to \( x_1 \). It is straightforward to extend this model to multiple tracers with MIDs \( x^1, x^2, \ldots, x^K \),

\[
x = \sum_{k=0}^{K} a_k x^k
\]

Provided that \( K \leq n \), we can still solve for the mixture coefficients \( a_k \).

An important generalization is to apply the above mixture model also in cases where some of the mixture components are metabolic products. For example, Rosenblatt et al. [1992] described a case where a doubly labeled leucine tracer is metabolized to a singly labeled one, and both are mixed with endogenous leucine (Figure 2). To solve this model, we must of course know the MID \( x^1 \) of the biologically synthesized \(^{13}\)C\(_1\)-leucine, as well as the tracer MID \( x^2 \). Since \( x^1 \) cannot be observed experimentally, Rosenblatt et al. [1992] suggested to instead use a theoretical distribution, which we will describe next.

2.2 The MID of labeled compounds

In this section, we describe the MID of an \( n \)-carbon compound that has incorporated \( k \) atoms from a pure \(^{13}\)C source and \( n - k \) atoms from a natural distribution, \( 0 \leq k \leq n \). Denote this MID \( \tau^{n,k} \). For the special case \( k = 0 \) we simply have \( \tau^{n,0} = \beta^n \), while for \( k = n \) we have the pure \(^{13}\)C\(_n\) MID \( \tau^{n,n} = e^{n,n} \). For \( 0 < k < n \), we assume that the \( n \)-carbon compound is formed by a condensation between a pure \(^{13}\)C\(_k\) compound with MID \( \tau^{k,k} \) and a naturally distributed compound with MID...
Figure 2: Mixture model adapted from Rosenblatt et al. [1992], with some simplifications. Here a $^{13}$C$_2$ leucine tracer with MID $x^2$ is metabolized to a $^{13}$C$_1$-leucine compound ($x^1$), and both are mixed with endogenous leucine ($x^0$), with mixture coefficients $a_k$.

Figure 3: (A) Condensation between a pure $^{13}$C$_2$ compound with MID $\tau^{2,2}$ and a C$_4$ compound with natural distribution $\tau^{4,0} = \beta^4$, yielding the MID $\tau^{6,2}$. (B) Example of step-wise condensations of pure $^{13}$C atoms with natural $^{13}$C atoms, resulting in the same MID $\tau^{6,2}$.

$\tau^{n-k,0}$ (see Figure 3A for an example). The resulting MID $\tau^{n,k}$ of the condensation product is then given by the convolution

$$\tau^{n,k} = \tau^{k,k} \otimes \tau^{n-k,0} = e^{k,k} \otimes \beta^{n-k}$$

where $x \otimes y$ is defined component-wise as

$$(x^1 \otimes x^2)_k = \sum_{i+j=k} x^1_i x^2_j, \quad 0 \leq k \leq m + n .$$

It is straightforward to show that this convolution yields the MID

$$\tau^{n,k} = \begin{pmatrix} 0^k \\ \beta^{n-k} \end{pmatrix}$$

where $0^k$ denotes the $k$-dimensional zero vector. That is, convoluting a natural MID with the unit vector $e^{k,k}$ will "shift" the MID by $k$ steps. For example, the labeled compounds in Figure 2 would have MIDs

$$x^1 = \tau^{6,1} = \begin{pmatrix} 0^5 \\ \beta^5 \end{pmatrix} \quad x^2 = \tau^{6,2} = \begin{pmatrix} 0^2 \\ \beta^4 \end{pmatrix}$$

This model based on condensations reflects how larger molecules are synthesized by anabolic pathways, which is indeed how most tracers are produced: for example, the MID $\tau^{6,2}$ might be produced by a condensation $\tau^{6,2} = \tau^{2,2} \otimes \tau^{4,0}$ (Figure 3A). Moreover, any sequence of condensations (in any order) that incorporate a total of $k$ $^{13}$C atoms from a pure $^{13}$C source and $n - k$ atoms from a natural carbon source will generate same MID $\tau^{n,k}$. This is because a condensation between any
two \( \tau \) distributions always yields another \( \tau \) distribution,

\[
\tau^{m,k} \otimes \tau^{n,l} = (e^{k,k} \otimes \beta^{m-k}) \otimes (e^{l,l} \otimes \beta^{n-l}) \\
= (e^{k,k} \otimes e^{l,l}) \otimes (\beta^{m-k} \otimes \beta^{n-l}) \\
= e^{k+l,k+l} \otimes \beta^{(m+n)-(k+l)} = \tau^{m+n,k+l}
\]

since the \( \otimes \) operator is commutative and associative, and it holds that \( \beta^m \otimes \beta^n = \beta^{m+n} \) for any \( m,n \). Applying this property repeatedly, we see that an MID \( \tau^{n,k} \) can be formed from 1-carbon building block as

\[
\tau^{n,k} = \tau^{1,1} \otimes \cdots \otimes \tau^{1,1} \otimes \tau^{1,0} \otimes \cdots \otimes \tau^{1,0} = \tau^{k,k} \otimes \tau^{n-k,0}.
\]

where the condensations can occur in any order. For example, Figure 3B shows one possible sequence of condensations giving rise to \( \tau^{6,2} = \tau^{2,2} \otimes \tau^{4,0} \).

With the class of distributions \( \tau^{n,k} \), we can now model all MIDs in the example of Figure 2 as \( x^0 = \tau^{6,0}, x^1 = \tau^{6,1} \) and \( x^2 = \tau^{6,2} \), and this generalizes to any mixture of compounds produced by condensations. However, the resulting mixture MID \( x \) in Figure 2 is not itself a \( \tau \) distribution.

### 2.3 From mixture models to MID correction

We now derive the general case of the MID correcting transform \( T^n \) from the distributions \( \tau^{n,k} \). Fernandez et al. [1996] described the case of an arbitrary \( n \)-carbon molecule formed as a mixture of \( n+1 \) compounds with distributions \( \tau^{n,0}, \tau^{n,1}, \ldots, \tau^{n,n} \), defined as above. We can write the resulting mixture model in matrix form as

\[
x = T^n a
\]

where \( a \) is again the vector of mixture coefficients, and \( T^n \) is the \( (n+1) \times (n+1) \) matrix

\[
T^n = (\tau^{n,0} \ \tau^{n,1} \ \cdots \ \tau^{n,n})
\]
whose columns are the MID vectors of the mixed compounds. Comparing with (2), we see that this is exactly the MID correction matrix (1), if the vector of mixture coefficients \(a\) can be interpreted as the corrected MID \(y\). To understand why this interpretation is valid, consider the following thought experiment (Figure 4).

For any metabolite with any MID \(x\), we can partition the population of molecules into \(n+1\) subpopulations, where subpopulation \(k\) consists of those molecules that have incorporated \(k\) \(^{13}\text{C}\) atoms from a pure \(^{13}\text{C}\) source and \(n-k\) atoms from a natural source, \(k = 0, 1, \ldots, n\) (Figure 4A). Then, the MID of subpopulation \(k\) is precisely \(\tau^{n,k}\), and \(a_k\) is the the fraction of molecules belonging to this sub-population, and we again have the mixture MID \(x = T^n a\). Now, imagine that we could replace all natural-derived \(^{13}\text{C}\) in these molecules with \(^{12}\text{C}\) (Figure 4B). Then the MID of subpopulation \(k\) would simply be the unit vector \(e^{n,k}\), and therefore the mixture MID becomes identical to the mixture coefficients, \(y = \sum_{k=0}^{n} a_k e^{n,k} = a\). Therefore, the vector \(a\) can be interpreted as the MID \(y\) that we would have obtained if natural \(^{13}\text{C}\) did not exist, that is, the corrected MID.

In summary, we have demonstrated that for any \(n\)-carbon compound formed by combining pure \(^{13}\text{C}\) carbon with natural-distributed carbon, the observed MID \(x\) and the corrected MID \(y\) are related as \(x = T^n y\). This model is valid for any product of a metabolic network whose substrates are either pure \(^{13}\text{C}\) \((e^{n,n})\) or natural \((\beta^n)\).

### 3 Properties of the transform \(T^n\)

In this section, we discuss some interesting properties of the transform \(T^n\). To gain a bit of intuition, Figure 5 visualizes the effects of \(T^n\) for the case \(n = 2\) at various points in the 2-simplex \(S_2\). (The \(n\)-simplex \(S_n\) is the set \(\{x \in \mathbb{R}^{n+1} : x_i \geq 0 \land \sum_{i=0}^{n} x_i = 1\}\), that is, the set of all possible \(n\)-carbon MIDs.) As with any linear operator, \(T^n\) maps the \(k\)'th unit vector \(e^{n,k}\) to the MID given by the \(k\)'th column of \(T^n\), \(T^n e^{n,k} = \tau^{n,k}, k = 0, \ldots, n\). A special case is \(k = n\) for which \(T^n e^{n,n} = \tau^{n,n} = e^{n,n}\).
This is the only eigenvector of \( T^n \) that lies within the simplex \( S_n \), reflecting the fact that a pure \( ^{13}\text{C} \) compound cannot have incorporated natural \( ^{13}\text{C} \). All other vectors \( y \in S_n \) are “pushed” towards \( e^{n,n} \) by \( T^n \) (Figure 5). Conversely, \( (T^n)^{-1} \) maps each MID \( \tau^{n,k} \) to the unit vector \( e^{n,k} \), and in particular maps the natural MID \( \tau^{n,0} = \beta^n \) to the unit vector \( e^{n,0} \). It is in this sense that \( (T^n)^{-1} \) ”removes” natural \( ^{13}\text{C} \). However, it should be noted that there is nothing about the corrected MID \( y = (T^n)^{-1}x \) that distinguishes it as inherently free of natural \( ^{13}\text{C} \): we could in principle observe the same MID \( y \) from another measurement. This is evident from the fact that \( (T^n)^{-1} \) is not a projection, as \( (T^n)^{-1}(T^n)^{-1}x \neq (T^n)^{-1}x \).

It is important to note that, while \( T^n \) is invertible, it is not surjective. It is clear that if \( y \in S_n \), then \( x = T^n y \) is also in \( S_n \), since \( T^n y \) is a convex combination of the columns of \( T^n \), all of which are in \( S_n \). But not all vectors \( x \in S_n \) can be written as \( x = T^n y \); that is, the image of \( T^n(S_n) \) does not cover all of \( S_n \) (Figure 5). For example, no unit vector \( e^{n,k} \) for \( k < n \) is in \( T^n(S_n) \). The part of \( S_n \) that lies outside \( T^n(S_n) \) are those MIDs cannot be corrected by \( (T^n)^{-1} \). These are MIDs that cannot be formed by condensing pure \( ^{13}\text{C} \) and natural-distributed carbon as described in Section 2.2. Generally, such MIDs vectors contain less \( ^{13}\text{C} \) than expected from the natural distribution. MID vectors of this kind can be generated by measurement errors, for example if a mass isotopomer peak is underestimated in mass spectrometry analysis. Attempting to apply \( (T^n)^{-1} \) to such an MID \( x \) will yield an \( y \) outside \( S_n \), and can result in negative elements of \( y \), as other have observed [Millard et al., 2012].

Regarding numeric stability of the transform, it is easy to show that the eigenvalues of \( T^n \) are \( \lambda_k = (1 - p)^k, k = 0, \ldots, n \), which gives the matrix condition number \( \lambda_0/\lambda_n = (1 - p)^{-n} \). Although this number increases exponentially with molecule size \( n \), since \( 1 - p \) is close to 1, it remains small for practically relevant values of \( n \). For example, with \( p = 0.01 \), we have \( (1 - p)^{-50} = 1.65 \), indicating that \( T^n \) is stable for most practical purposes.

4 Flux analysis on corrected MIDs

We now turn to the question of how MID correction might affect metabolic flux analysis. We will restrict here ourselves to the case of metabolic and isotopic steady-state. A metabolic network acts on a set of substrates with MIDs \( \mathbf{X}_{in} \) to generate products with MIDs \( \mathbf{X} \). Here we use bold \( \mathbf{X} \)
to indicate a set of MID vectors of various lengths. Each MID in $X_{in}$ may be any $x \in T^n(S_n)$, including $^{13}$C tracers and endogenous compounds. At steady-state, given a flux vector $v$, product MIDs $X$ are uniquely determined by a nonlinear function $X = f_v(X_{in})$ of the substrate MIDs [Wiechert and de Graaf, 1997]. We will refer to $f_v$ as the network function. Write $X = T Y$ to indicate the transform $T$ that maps each $n$-carbon MID as $x = T^n y$. For flux analysis on corrected MIDs to be valid, it must hold for any flux vector $v$ that

$$Y = f_v(Y_{in}) \iff X = f_v(X_{in})$$

(4)

This criterion says that, if the observed product MIDs $X$ are generated from the substrates $X_{in}$ by the network, then the corrected product MIDs $Y$ will be generated by the same network from the corrected substrate MIDs $Y_{in}$, and vice versa. In other words, a flux vector $v$ will fit the observed MIDs if and only if it fits the corrected MIDs, so metabolic flux analysis will give the same results in both cases. For (4) to hold true, it will be essential that we can exchange order between $f_v$ and $T$. The following theorem is key.

**Theorem 1.** For any metabolic network and flux vector $v$, the network function $f_v$ satisfies

$$T f_v(Y_{in}) = f_v(T Y_{in}) .$$

Theorem 1 implies the equivalence (4) because, if $Y = f_v(Y_{in})$, then $X = TY = T f_v(Y_{in}) = f_v(T Y_{in}) = f_v(X_{in})$. Conversely, if $X = f_v(X_{in})$, then $Y = T^{-1} f_v(X_{in}) = f_v(T^{-1} X_{in}) = f_v(Y_{in})$ as well. An illustration of the theorem for an example metabolic network is shown in Figure 6. We will first prove Theorem 1 for the special case of a compartmental network, and then prove the general case by representing a metabolic network as a cascade of compartmental networks connected by condensations, as described previously [Wiechert et al., 2001, Antoniewicz et al., 2007].

### 4.1 Compartmental networks

A compartmental network is a directed graph over $M$ substrates and $N$ products, all of which are $n$-carbon metabolites. At steady-state, for a given flux vector $v$, the compartmental network satisfies the mass balance equations

$$A(v) X + B(v) X_{in} = 0$$

(5)

where $X$ is an $N \times (n + 1)$ matrix whose rows are the product MIDs, and $X_{in}$ is an $M \times (n + 1)$ matrix of substrate MIDs. The $N \times N$ matrix $A(v)$ describes fluxes between products, and is always invertible [Anderson, 1979], while the $N \times M$ matrix $B(v)$ describes fluxes from substrates to products. In this case, the network function $f_v$ is found by solving the linear system (5), $X = f_v(X_{in}) = -A(v)^{-1} B(v) X_{in}$, and is easy to demonstrate that $T$ and $f_v$ can be exchanged.

**Lemma 1.** Let $f_v$ be the network function of a compartmental network over $n$-carbon metabolites. Then it holds that

$$T f_v(Y_{in}) = f_v(T Y_{in}) .$$

**Proof.** In this case, we can write the transform $T$ as a matrix multiplication $X = (T^n Y^\top)^\top = Y(T^n)^\top$. Since $T^n$ and $f_v$ are both linear, we obtain

$$T f_v(Y_{in}) = f_v(Y_{in})(T^n)^\top = -A^{-1}B Y_{in}(T^n)^\top = f_v \left( Y_{in}(T^n)^\top \right) = f_v(T Y_{in}) .$$

\[\square\]
4.2 General metabolic networks

Any metabolic network can be written as a cascade of compartmental sub-networks, for molecule sizes \( s = 1, 2, \ldots, S \), connected by condensations that merge two smaller molecules into a larger one [Wiechert et al., 2001, Antoniewicz et al., 2007]. In this model, "molecules" need not be physical compounds, but can also represent various moieties of such compounds; however, this makes no difference for our purposes. At steady-state, each compartmental subnetwork satisfies the balance equation

\[
A^s(v)X^s + B^s(v)X^s_{\text{in}} + C^s(v)X^s_{\text{cond}} = 0, \quad s = 1, \ldots, S
\]

where \( A^s(v) \), \( B^s(v) \) and \( C^s(v) \) are matrices describing the subnetwork structure, \( X^s \) and \( X^s_{\text{in}} \) are product and substrate MIDs matrices as before, and \( X^s_{\text{cond}} \) is an additional MID matrix formed by condensations \( x^s = x^m \otimes x^n \), where \( x^m \) and \( x^n \) are MIDs from \( X^m \) and \( X^n \), for some \( m, n \) such that \( m + n = s \). Thus, each \( X^s_{\text{cond}} \) is a function \( X^s_{\text{cond}}(X^1, \ldots, X^{s-1}) \) of the product MIDs from smaller size subnetworks. Let \( f^s_v(X^s_{\text{in}}, X^s_{\text{cond}}) \) denote the network function for compartmental subnetwork \( s \). Viewing \( X^s_{\text{cond}} \) as a known constant, \( f^v \) is again a linear function, given by solving (6). The entire metabolic network then has a network function \( f^{1:S}_v(X^1_{\text{in}}, \ldots, X^S_{\text{in}}) \) defined as

\[
X = (X^1, \ldots, X^S) = (f^1_v(X^1_{\text{in}}, X^1_{\text{cond}}), f^2_v(X^2_{\text{in}}, X^2_{\text{cond}}(X^1_{\text{in}})), \ldots, f^S_v(X^S_{\text{in}}, X^S_{\text{cond}}(X^1_{\text{in}}, \ldots, X^{S-1}_{\text{in}}))) = f^1:S_v(X^1_{\text{in}}, \ldots, X^S_{\text{in}})
\]

Because each \( X^s_{\text{cond}} \) is a formed by convolution of the products from smaller size subnetworks, the proof of Theorem 1 will require that we also can exchange the convolution operator \( \otimes \) with the transform \( T \), so that \( TX^s_{\text{cond}}(X^1, \ldots, X^{s-1}) = X^s_{\text{cond}}(TX^1, \ldots, TX^{s-1}) \). This is guaranteed by the following lemma.

**Lemma 2.** For any two MID vectors \( x^m \in S_m \) and \( x^n \in S_n \), it holds that

\[
(T^m x^m) \otimes (T^n x^n) = T^{m+n}(x^m \otimes x^n)
\]

A proof is given in Appendix A. We are now ready to prove Theorem 1 for the general case.

**Proof of Theorem 1.** We will proceed by induction over the size \( t \) of the largest subnetwork. For \( t = 1 \) the theorem is true by Lemma 1. For \( t > 1 \), we assume that the theorem holds for the network formed by subnetworks \( s = 1, \ldots, t-1 \), that is, \( T f^{1:t-1}_v(Y^1_{\text{in}}, \ldots, Y^{t-1}_{\text{in}}) = f^{1:t-1}_v(TY^1_{\text{in}}, \ldots, TY^{t-1}_{\text{in}}) \). For the size \( t \) compartmental subnetwork, we have

\[
T f^t_v(Y^t_{\text{in}}, Y^t_{\text{cond}}) = f^t_v(TY^t_{\text{in}}, TY^t_{\text{cond}})
\]

by Lemma 1, and

\[
TY^t_{\text{cond}}(Y^1, \ldots, Y^{t-1}) = Y^t_{\text{cond}}(TY^1, \ldots, TY^{t-1})
\]

by Lemma 2. Together with the induction assumption, this gives

\[
T f^{1:t}_v(Y^1_{\text{in}}, \ldots, Y^t_{\text{in}}) = (T f^{1:t-1}_v(Y^1_{\text{in}}, \ldots, Y^{t-1}_{\text{in}}), T f^t_v(Y^t_{\text{in}}, Y^t_{\text{cond}})) = (f^{1:t-1}_v(TY^1_{\text{in}}, \ldots, TY^{t-1}_{\text{in}}), f^t_v(TY^t_{\text{in}}, Y^t_{\text{cond}}(TY^1, \ldots, TY^{t-1}))) = f^{1:t}_v(TY^1_{\text{in}}, \ldots, TY^t_{\text{in}})
\]

which completes the induction step. \( \square \)
In this paper, we have provided a self-contained derivation of the linear transform $T^n$ commonly used for isotope correction, and explored the effects of $T^n$ for metabolic flux analysis. We hope that this contribution will help clarify the role of isotope correction in analysis of mass spectrometry data from $^{13}$C tracing experiments. It should be clear from our derivation that $T^n$ is not an empirical error correction (as the name might suggest), but a rather a transformation that yields a particularly simple representation, the theoretical "corrected" MID $y$, which by definition contains no natural-derived $^{13}$C. At first glance, it might seem remarkable that a single linear transform exists that can "remove" $^{13}$C from any observed metabolite, with any isotopomer distribution whatsoever. This is possible because we have assumed a particular underlying model generating labeled compounds, namely that $^{13}$C enters metabolites by condensations of pure $^{13}$C atoms with natural $^{13}$C distributed atoms (section 2.2). This process always yields MIDs of the type $\tau^{n,k}$ for an $n$-carbon compound with $k$ heavy atoms, and the argument in section 2.3 shows that $(T^n)^{-1}$ maps any mixture of $\tau^{n,k}$ MIDs to a corrected MID $y$. However, it is important to note that $(T^n)^{-1}$ is not applicable to MID vectors that cannot be generated by the process we assumed; in particular, MIDs with less $^{13}$C enrichment than the natural distribution are not feasible. This can cause problems in cases where the measured mass isotopomer fractions are biased or uncertain.

In our derivation of the class of distributions $\tau^{n,k}$, we only considered the case of pure $^{13}$C tracers. In practise, tracers always have some degree of $^{12}$C impurities. To handle this complication, the $^{13}$C source can be considered to emit $^{13}$C with a propability $q < 1$ (where $q$ is typically around 0.98 – 0.99), and the pure $^{13}$C distribution $e^{k,k}$ is then replaced with another binomial, say $a^n = \text{Bin}(n,q)$. We would then obtain a different labeled compound distribution $\tau^{n,k} = a^l \otimes b^{n-k}$, and can construct a new linear transform $T^n$ based on this distribution. Importantly, equation (3) still holds for such $\tau$ distributions, and therefore the proof of Lemma 2 remains valid. Hence, our main result is still applicable in this more general case.

One practical advantage of isotope correction is that the vectors $y$ are typically sparse, since only atoms that derive from the pure $^{13}$C source (that is, the tracer) can yield nonzeros in $y$. Hence, in the typical isotope tracing experiment, most mass isotopomer fractions $y_i$ will be zero. This sparsity is intuitively exploited by researchers creating metabolic network models, by manually excluding moieties that cannot possibly be reached by tracer-derived atoms, such as cofactors and essential amino acids. However, this process could be performed automatically for any given metabolic network, by discarding any atoms that cannot be reached from tracer $^{13}$C. For example, in a previously published model of one-carbon metabolism with U-$^{13}$C5-methionine as tracer [Nilsson et al., 2017], a total of 341 atoms had to be included to simulate the measured MIDs, but only 65 atoms are needed to simulate the corresponding corrected MIDs.

Although it is clear that the isotope-correcting transform is widely applicable, it is not obvious that metabolic flux analysis (MFA) on a set of corrected MIDs $Y$ is equivalent to MFA on the corresponding set of observed MIDs $X$. Our theorem 1 provides a theoretical foundation for using MFA on corrected MIDs, by guaranteeing that any set of steady-state isotopomer balances (6) hold for $Y$ if and only if they hold for $X$. In other words, isotopomer balances for $Y$ and $X$ are compatible with precisely the same set of flux vectors $v$. This equivalence holds because $T^n$ is linear and commutes convolution of MIDs (Lemma 2). However, an important limitation is that Theorem 1 applies only when the isotopomer balances hold exactly. In practice, where models are inaccurate and observed MIDs $X$ are afflicted by measurement errors, there is generally no vector $v$ that that exactly satisfies $X = f_v(X_{in})$. Instead, we must seek a $v$ that minimizes the error.
\[ \| X = f_v(X_{in}) \|, \text{ where } \| \cdot \| \text{ typically is the } L_2 \text{ (Euclidean) norm. This will in general not yield the same solution as minimizing } \| Y = f_v(Y_{in}) \|, \text{ since the } L_2 \text{ norm will measure errors differently in the two representations. Consequently, both point estimates and confidence intervals on fluxes will generally differ between the two representations. This issue should be further investigated, and we would recommend that flux solutions from MFA on corrected data } Y \text{ should be confirmed to also fit the uncorrected data } X. \text{ Interestingly, this problem could also be viewed as an indication that the } L_2 \text{ norm is not appropriate for this setting, and it could be argued that a suitable norm should generate identical solutions in both representations. Whether there exists a norm on the } n \text{-simplex } S_n \text{ satisfying this criterion, for example one as the measures described by Aitchison [1982], is an interesting topic for future studies. Finally, we have only treated metabolic networks at metabolic and isotopic steady-state. It seems that the proof strategy employed here should also extend to isotopic non-steady state (which involves a cascade of differential equations of a similar structure), but a rigorous proof for this scenario clearly requires further work.}

Acknowledgements

The author wishes to thank Dr. Firas Midani for helpful suggestions on the manuscript. This work was funded by the Foundation for Strategic Research grant no.ITM17-0245 and Karolinska Institutet.

References

John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B. Methodological*, 44(2):139–177, 1982.

David H Anderson. Iterative inversion of single exit compartmental matrices. *Comput Biol Med*, 9:317–330, 1979.

Maciek R Antoniewicz, Joanne K Kelleher, and Gregory Stephanopoulos. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic engineering*, 9 (1):68–86, 2007.

Michael Berglund and Michael E. Wieser. Isotopic compositions of the elements 2009 (IUPAC Technical Report). *Pure and Applied Chemistry*, 83(2):397–410, 2011.

John. I. Brauman. Least squares analysis and simplification of multi-isotope mass spectra. *Analytical Chemistry*, 38(4):607–610, 1966.

Joerg M Buescher, Maciek R Antoniewicz, Laszlo G Boros, Shawn C Burgess, Henri Brunengraber, Clary B Clish, Ralph J DeBerardinis, Olivier Feron, Christian Frezza, Bart Ghesquiere, Eyal Gottlieb, Karsten Hiller, Russell G Jones, Jurre J Kamphorst, Richard G Kibbey, Alec C Kimelman, Jason W Locasale, Sophia Y Lunt, Oliver D K Maddocks, Craig Malloy, Christian M Metallo, Emmanuelle J Meuillet, Joshua Munger, Katharina Nöhl, Joshua D Rabinowitz, Markus Ralser, Uwe Sauer, Gregory Stephanopoulos, Julie St-Pierre, Daniel a Tennant, Christoph Wittmann, Matthew G Vander Heiden, Alexei Vazquez, Karen Vousden, Jamey D Young, Nicola Zamboni, and Sarah Maria Fendt. A roadmap for interpreting $^{13}$C metabolite labeling patterns from cells. *Current Opinion in Biotechnology*, 34:189–201, 2015.
C. Cobelli, G. Toffolo, and D. M. Foster. Tracer-to-tracee ratio for analysis of stable isotope tracer data: link with radioactive kinetic formalism. *American Journal of Physiology-Endocrinology and Metabolism*, 262(6):E968–E975, 1992.

Charles A. Fernandez, Christine Des Rosiers, Stephen F. Previs, France David, and Henri Brunengraber. Correction of $^{13}$C Mass Isotopomer Distributions for Natural Stable Isotope Abundance. *Journal of Mass Spectrometry*, 31(3):255–262, 1996.

Christian Jungreuthmayer, Stefan Neubauer, Teresa Mairinger, Jürgen Zanghellini, and Stephan Hann. ICT: isotope correction toolbox. *Bioinformatics*, 32(1):154–156, 2015.

Firas S. Midani, Michelle L. Wynn, and Santiago Schnell. The importance of accurately correcting for the natural abundance of stable isotopes. *Analytical Biochemistry*, 520:27–43, 2017.

P. Millard, F. Letisse, S. Sokol, and J.-C. Portais. Isocor: correcting ms data in isotope labeling experiments. *Bioinformatics*, 28(9):1294–1296, 2012.

Pierre Millard, Baudoin Delâ©pine, Matthieu Guionnet, Maud Heuillet, Floriant Bellvert, and Fabien LÀ©tisse. IsoCor: isotope correction for high-resolution MS labeling experiments. *Bioinformatics*, 35(21):4484–4487, 2019.

Roland Nilsson, Irena Roci, Jeramie Watrous, and Mohit Jain. Estimation of flux ratios without uptake or release data: Application to serine and methionine metabolism. *Metabolic Engineering*, 43:137 – 146, 2017.

J. Rosenblatt, D. Chinkes, M. Wolfe, and R. R. Wolfe. Stable isotope tracer analysis by GC-MS, including quantification of isotopomer effects. *American Journal of Physiology-Endocrinology and Metabolism*, 263(3):E584–E596, 1992.

Xiaoyang Su, Wenyun Lu, and Joshua D. Rabinowitz. Metabolite spectral accuracy on orbitraps. *Analytical Chemistry*, 89(11):5940–5948, 2017.

W Wiechert and A A de Graaf. Bidirectional reaction steps in metabolic networks: I. modeling and simulation of carbon isotope labeling experiments. *Biotechnology and bioengineering*, 55(1):101–17, July 1997.

W Wiechert, M Möllney, S Petersen, and a a de Graaf. A universal framework for 13c metabolic flux analysis. *Metabolic engineering*, 3(3):265–83, 2001.
A Proof of Lemma 2

Proof of Lemma 2. For the case of basis vectors $x = e^{m,k}$ and $y = e^{n,l}$, we already know that

$$(T^m e^{m,k}) \otimes (T^n e^{n,l}) = \tau^m \otimes \tau^n = \tau^{m+n,k+l} = T^{m+n} e^{m+n,k+l} = T^{m+n} (e^{m,k} \otimes e^{n,l})$$

from equation (3). For arbitrary vectors $x \in S_m$ and $y \in S_n$, write $x = \sum_{i=0}^{m} x_i e^{m,i}$ and $y = \sum_{j=0}^{n} y_j e^{n,j}$. Then

$$(T^m x) \otimes (T^n y) = \left( \sum_{i=1}^{m} x_i T^m e^{m,i} \right) \otimes \left( \sum_{j=1}^{n} y_j T^n e^{n,j} \right)
= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j T^m e^{m,i} \otimes T^n e^{n,j}
= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j T^{m+n} (e^{m,i} \otimes e^{n,j})
= T^{m+n} \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j e^{m,i} \otimes e^{n,j}
= T^{m+n} (x \otimes y)$$

where in step * we have used the fact that the $\otimes$ operation distributes over vector addition, that is, $a \otimes (b+c) = a \otimes b + a \otimes c$ for any vectors $a, b, c$. \qed