Metabolic Flux Analysis in Isotope Labeling Experiments Using the Adjoint Approach

Stéphane Mottelet, Gil Gaullier, and Georges Sadaka

Abstract—Comprehension of metabolic pathways is considerably enhanced by metabolic flux analysis (MFA-ILE) in isotope labeling experiments. The balance equations are given by hundreds of algebraic (stationary MFA) or ordinary differential equations (nonstationary MFA), and reducing the number of operations is therefore a crucial part of reducing the computation cost. The main bottleneck for deterministic algorithms is the computation of derivatives, particularly for nonstationary MFA. In this article, we explain how the overall identification process may be speeded up by using the adjoint approach to compute the gradient of the residual sum of squares. The proposed approach shows significant improvements in terms of complexity and computation time when it is compared with the usual (direct) approach. Numerical results are obtained for the central metabolic pathways of *Escherichia coli* and are validated against reference software in the stationary case. The methods and algorithms described in this paper are included in the `sysmetab` software package distributed under an Open Source license at http://forge.scilab.org/index.php/p/sysmetab/.

Index Terms—Metabolic engineering, isotopically non-stationary, inverse problem, parameter identification, non-linear optimization, open source, software, sysmetab, FLUXML

1 Introduction

Nonstationary MFA can contribute significantly to the comprehension, both qualitative and quantitative, of a metabolic network [1], but before the creation of the Elementary Metabolite Units (EMU) framework [2], [3] the high computation cost of nonstationary MFA prevented the development of efficient software. However, topological network reduction methods of the sort that EMU makes possible are not the only way of curbing the computation cost of nonstationary MFA. Given that both stationary and nonstationary MFA belong to the class of general inverse problems where measurements depend on parameters through a state equation, there is a technique, namely the adjoint approach [4], [5], [6], that may be used to greatly improve the computation of different kinds of derivatives.

The main contribution of this paper is demonstrating that independently of the topological network reduction, the adjoint approach considerably speeds up the computation of the gradient residual sum of squares (RSS), meaning that the entire process of estimating unknown parameters is accelerated significantly. We show that in the (usual) direct approach the sum of the times required for the state and gradient calculations is dominated by the computation time of the gradient itself, whereas in the adjoint approach the two computation times are of the same order. While the direct approach has a cost that is unavoidably proportional to the number p of unknown parameters (fluxes and pool sizes), in the case of the adjoint approach the cost of the gradient computation is independent of p. The algorithms described below are implemented in the accompanying Open Source software `sysmetab`, which can handle stationary and nonstationary MFA.

As theoretical results are asymptotical and can be degraded by practical implementation bottlenecks, we perform a time analysis for both gradient computations on the central metabolism of *E. Coli*, in which we compare overall computation times for the direct and adjoint approaches with respect to labeling state sizes: reduced versus full cumomer set and stationary versus nonstationary data. We contrast estimated flux values for the stationary data with the values obtained from the reference softwares `13CFLux2` [7] and `influx_s` [8], which do not implement the adjoint approach. We detail the distribution of running time on the network mentioned above and on a larger, poorly defined version of this network. We show that in this particular context, when multiple flux estimations from different perturbed data are made repeatedly (i.e., the Monte Carlo method is applied), `sysmetab` is competitive with other software. For parameter estimations using time-dependent labeling data generated by stationary flux estimations and realistic pool size values, the ratio of running times from one approach to the other exceeds 20.

The paper is organized as follows. In Section 2, we introduce the mathematical structure of ^{13}C MFA. Mathematical aspects of the algorithms implemented in `sysmetab` are given in Sections 3 and 4 respectively for stationary and nonstationary MFA, and their complexity is compared. Implementation details are described in Section 5. Section 6 presents numerical results obtained by `sysmetab`.

2 Optimization Model of MFA

The aim of ^{13}C metabolic flux analysis is to determine fluxes v, pool sizes m (in the nonstationary case), expressed as a vector of p parameters $\theta = (v, m)$, such that experimental labeling data y and other measurements w, e.g., input/output fluxes and pool sizes, are best fitted to their simulated values. This means minimizing the objective function

$$J(\theta) = \frac{1}{2}\|g(x(\theta), \theta) - y\|_{\mathcal{Y}}^2 + \frac{1}{2}\|h(\theta) - w\|_{\mathcal{W}}^2, \tag{1}$$

where g and h are given functions and $x(\theta)$ denotes the (possibly time-dependent) n labeling states of the metabolic network, implicitly defined by the state equation

$$f(x, \theta) = 0. \tag{2}$$

This equation takes the form of a system of algebraic or ordinary differential equations (ODE), respectively for stationary or nonstationary metabolic flux analysis. The notations $\|\cdot\|_{\mathcal{Y}}$ and $\|\cdot\|_{\mathcal{W}}$ are used for the norms in the space of labeling and non-labeling measurements, denoted by $\mathcal{Y}$ and $\mathcal{W}$ respectively. These norms (may) model variations due to experimental noise by introducing weighting covariance matrices. In both the stationary and nonstationary cases the spaces $\mathcal{Y}$ and $\mathcal{W}$ are finite-dimensional normed vector spaces, i.e., y and w can be considered as vectors containing real measured data.

Below, Θ denotes the space of admissible parameters, i.e., the space describing realistic physical parameters comprising constraints on fluxes, such as the stoichiometric equations for fluxes. In order to find

$$\hat{\theta} = \arg\min_{\theta \in \Theta} J(\theta), \tag{3}$$

different optimization methods can be considered. Finding $\hat{\theta}$ can be performed by determining the solution of successive linearized problems, as the in Gauss-Newton method (GN) with a first-order Taylor expansion of $g(x(\theta))$ with respect to θ by the computation of the sensitivity matrix $x'(\theta)$. In the nonstationary case this $n \times p$ matrix is time-dependent and satisfies a system of differential equations, which means that it is very expensive to compute.

- *S. Mottelet is with the Department EA4297 TIMR, Sorbonne University, Université de technologie de Compiègne, rue du docteur Schweitzer, sF-60203, Compiègne, France. E-mail: stephane.mottelet@utc.fr.*
- *G. Gaullier is with the Department EA2222 LMAC, Sorbonne University, Université de technologie de Compiègne, France. E-mail: gil.gaullier@utc.fr.*
- *G. Sadaka is with the Department CNRS, UMR7352 LAMFA, Université de Picardie Jules Verne, France. E-mail: georges.sadaka@u-picardie.fr.*

Manuscript received 31 July 2015; revised 19 Jan. 2016; accepted 10 Mar. 2016. Date of publication 21 Mar. 2016; date of current version 22 Mar. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2016.2544299

In contrast, if the minimization is considered as a general nonlinear optimization problem, only the gradient of J needs to be computed, as in the Broyden-Fletcher-Goldfarb-Shanno (BFGS) and Sequential Quadratic Problem (SQP) methods. In the 1970s the adjoint state method was developed to compute the gradient in the situation frequently encountered in control theory where the function to minimize depends indirectly on parameters through state variables, with a minimal cost [4], [5], [6]. This cost is asymptotically independent of the number of parameters p, and is equivalent to the cost of a single-state equation, while using the sensitivity matrix generates a cost proportional to p.

3 Stationary MFA

The cumomer fraction variables $x = (x_1, \ldots, x_N)$ model introduced in [9] means that the structure of balance equations in stationary MFA is given by a cascade of linear algebraic equations. If x_k denotes the vector of cumomer fractions of weight k , $x_k^{\leftarrow}$ the vector of input cumomers of weight k, and $x_{<k}$, $x_{\leq k}$ the sequences $(x_i)_{i<k}$ and $(x_i)_{i\leq k}$, respectively, then the state equation (2) can be written as

$$f_k(v, x_{\leq k}, x_k^{\leftarrow}) = 0, \quad 1 \leq k \leq N, \tag{4}$$

where the function f_k is affine with respect to x_k as

$$f_k(v, x_{\leq k}, x_{\leq k}^{\leftarrow}) = A_k(v)x_k + b_k(v, x_{<k}, x_{\leq k}^{\leftarrow}),$$

the matrix A_k and vector b_k being determined by the structure of the metabolic network under consideration. From the forward cascade structure of (4), we successively obtain $x_1, x_2, \ldots, x_N$ by solving at each step a system of linear equations. For the sake of conciseness we do not consider the terms which explicitly depend on θ in (1), but only labeling measurements. In this case, measurements linearly depend on the state variable, so that the functional to minimize is of the form

$$J(v) = \frac{1}{2}\|Cx(v) - y\|^2,$$

for a given matrix $C = (C_1, \ldots, C_N)$ depending on the measurement model.

3.1 Direct Approach

For a given flux vector v, we first need to solve the state equation, which provides the cumomer vector $x(v)$. The sequence of derivatives $(x'_k)_{1\leq k\leq N}$ is then computed by solving each step of the forward cascade (the sum in the second row vanishes for $k = 1$)

$$\begin{aligned} A_k(v)x'_k = &-\partial_v f_k(v, x_{\leq k}, x_{\leq k}^{\leftarrow}) \\ &+ \sum_{i<k} \partial_{x_i} b_k(v, x_{<k}, x_{\leq k}^{\leftarrow})x'_i, \ k \geq 1, \end{aligned} \tag{5}$$

which is obtained by implicit differentiation of (4) with respect to v. Finally, the gradient of J is computed as

$$\nabla J(v) = \sum_{k=1}^{N} x'_k(v)^{\top} C_k^{\top}(Cx(v) - y), \tag{6}$$

where the symbol $^{\top}$ denotes vector and matrix transposition.

3.2 Adjoint Approach

The adjoint approach is based on the Lagrangian functional

$$\mathcal{L}(x, v, \lambda) = \frac{1}{2}\|Cx - y\|^2 + \sum_{k=1}^{N} \lambda_k^{\top} f_k(v, x_{\leq k}, x_{\leq k}^{\leftarrow}),$$

where $\lambda = (\lambda_k)_{1\leq k\leq N}$ is the sequence of adjoint variables. Expanding the adjoint equation $\partial_x \mathcal{L}(x, v, \lambda) = 0$ [6] leads to the backward cascade (the sum vanishes for $k = N$)

$$\begin{aligned} A_k^{\top}(v)\lambda_k = &C_k^{\top}(Cx - y) \\ &- \sum_{i>k} \partial_{x_k} b_i(v, x_{<i}, x_{\leq i}^{\leftarrow})^{\top}\lambda_i, \ k \leq N. \end{aligned} \tag{7}$$

For a given v, equations (7) are solved backwards, once the cumomer vector $x(v)$ has been computed from (4). This procedure provides successively $\lambda_N, \lambda_{N-1}, \ldots, \lambda_1$. The gradient ∇J can then be obtained by the solutions of the N previous linear systems as follows

$$\nabla J(v) = \sum_{k=1}^{N} \partial_v f_k(v, x_{\leq k}, x_{\leq k}^{\leftarrow})^{\top}\lambda_k. \tag{8}$$

The adjoint approach can also be used to compute the derivative of any function of x with respect to v without requiring $x'(v)$ to be computed. One example of this would be the output sensitivity matrix $S(v) = Cx'(v)$, which may be used in first order sensitivity analysis [10] and to compute confidence intervals for the estimated parameters. This matrix is also used at each iteration of the optimization method to compute the descent direction when a Gauss-Newton type algorithm is used [8]). In both usages, the adjoint approach allows $S(v)$ to be computed as a fraction n_y/n of the usual cost, where n_y is the number of measurements and n the size of cumomer vector x.

The adjoint approach can also be used to compute the derivative of any function of x with respect to v without requiring $x'(v)$ to be computed. One example of this would be the output sensitivity matrix $S(v) = Cx'(v)$, which may be used t.

3.3 Computational Comparison

Both approaches involve the computation of the state $x(v)$ in the computation of $\nabla J(v)$. To achieve this, the matrices $A_k(v)$ need to be factorized and the derivatives $\partial_v f_k$ and $\partial_{x_i} b_k$ computed for $i < k$. The crucial difference between the two approaches is to be found in the two remaining equations: for a given k, equation (5) involves solving a linear system whose right-hand side is a matrix with p columns, whereas in equation (7) the right-hand side is a (single) column vector. A subsequent analysis of equations (6) and (8) shows that even though the elementary operations are identical, the computation cost of equation (8) differs from (6) in that $\partial_v f_k$ is sparse, unlike x'_k which is full. While the difference remains negligible for small metabolic networks, our numerical results show that the larger the network, the more significant the reduction in computation cost.

We show in the following section that for nonstationary MFA the improvements obtained from the adjoint approach are of the same magnitude. But, since the problem is time-dependent, the overall computation cost increases considerably. In order for the problem to become tractable, the adjoint approach is essential.

4 Nonstationary MFA

In nonstationary MFA, the balance equations are modeled by a cascade of ordinary differential equations given for $1 \leq k \leq N$ by

$$\begin{cases} x_k(0) = 0, \\ X_k(m)\dfrac{d}{dt}x_k(t) = f_k(v, x(t), x_k^{\leftarrow}), \ t \in]0, T]. \end{cases} \tag{9}$$

To simplify the presentation we will consider that the pool sizes are known. We consider that labeling measurements y_j are made at times $\tau_1 < \cdots < \tau_M$, where $0 < \tau_1$ and $\tau_M < T$, so that the functional to minimize is given by

$$J(v) = \frac{1}{2}\sum_{j=1}^{M} \|Cx(\tau_j; v) - y_j\|^2,$$

where $x(\tau_j; v)$ is the solution of (9) at time τ_j.

Since the cumomer fractions x_k are time-dependent, the space X of the labeling states is no longer a finite-dimensional vector space as in stationary MFA, but an infinite-dimensional functional space $L^2([0;T])$ given by square-integrable functions. The inner product $(\cdot,\cdot)_X$ in X is the usual inner product in $L^2([0;T])$, i.e., $(\phi,\psi)_X = \int_0^T \phi(t)^\top \psi(t) dt$.

4.1 Direct Approach

As in the stationary case, the direct approach is based on the calculation of the partial derivative x'_k with respect to the flux v. When the sequence $(x'_k)_{1\le k\le N}$ is time-dependent, this approach involves a cascade of ordinary differential equations

$$\begin{aligned} X_k(m)\frac{d}{dt}x'_k = \partial_v f_k(v, x_{\le k}, x^{\leftarrow}_{\le k}) + A_k(v)x'_k \\ + \sum_{i<k} \partial_{x_i} b_k(v, x_{\le i}, x^{\leftarrow}_{\le i})x'_i, \end{aligned} \tag{10}$$

for $1 \le k \le N$, obtained by the differentiation of (9) with respect to the flux parameter v. In (10), the sum in the second row vanishes for $k = 1$, and $\frac{d}{dt}x'_k$ denotes the time derivative of x'_k. The initial condition related to (10) is given, for all k such that $1 \le k \le N$, by $x'_k(0) = 0$. The gradient of J is expressed as

$$\nabla J(v) = \sum_{j=1}^{M}\sum_{k=1}^{N} x'_k(\tau_j; v)^\top C_k^\top (Cx(\tau_j; v) - y_j). \tag{11}$$

4.2 Adjoint Approach

In the adjoint approach, the N ODEs (9) are incorporated into a Lagrangian functional as follows

$$\begin{aligned} \mathcal{L}(x,p,\lambda) = \frac{1}{2}\sum_{j=1}^{M}\left\|Cx(\tau_j) - y_j\right\|^2 \\ + \sum_{k=1}^{N}\int_0^T \lambda_k^\top(t)\left(f_k(v, x_{\le k}(t), x^{\leftarrow}_{\le k}) - X_k(m)\frac{d}{dt}x_k(t)\right)dt. \end{aligned}$$

The adjoint equation is obtained by expanding $\partial_x \mathcal{L}(x,p,\lambda) = 0$, which leads to the following backwards-in-time ODE cascade

$$\begin{aligned} X_k(m)\frac{d}{dt}\lambda_k = A_k(v)^\top \lambda_k \\ - \sum_{i>k} \partial_{x_k} b_i(v, x_{<i}, x^{\leftarrow}_{\le i})^\top \lambda_i, \end{aligned} \tag{12}$$

for $1 \le k \le N$, where the second row vanishes for $k = N$ and the equalities read for $t \in [0, T[\backslash\{\tau_j\}_{1\le j\le M}$. The final condition is given for each k by $\lambda_k(T) = 0$ and the following jump conditions

$$X_k(m)[\lambda_k(\tau_j^+) - \lambda_k(\tau_j^-)] = C_k^\top(Cx(\tau_j) - y_j), \tag{13}$$

occur for $j = 1 \dots M$. From (12), (13) we obtain the functions $\lambda_k(t)$ which enter into the gradient of J (with respect to v) as follows

$$\nabla J(v) = \sum_{k=1}^{N}\int_0^T \partial_v f_k(v, x_{\le k}(t), x^{\leftarrow}_{\le k})^\top \lambda_k \, dt. \tag{14}$$

4.3 Discrete Time Schemes

Both the direct and the adjoint approaches in nonstationary MFA require state $x(t;v)$ to be computed, and this can be done efficiently using discrete schemes. From a time grid $t_n = nh$, with $h = T/N_T$ ($N_T + 1$ denoting the total number of points sampling the interval $[0;T]$), discrete schemes are based on approximating the integral of the right-hand side of (9) over $[t_n; t_{n+1}]$. For the sake of simplicity we consider the implicit Euler scheme, but a similar approach can be used with schemes having a greater order, while also taking into account the possible stiffness of the state equation, such as the implicit trapezoidal rule [11] or implicit Runge-Kutta schemes [12].

Denoting the approximations $x_k^n \simeq x_k(t_n)$ and $x^n_{\le k} \simeq x_{\le k}(t_n)$, the implicit Euler scheme gives the following discrete version of (9)

$$\begin{cases} x_k^0 = 0, \\ X_k(m)\big(x_k^{n+1} - x_k^n\big) = h f_k(v, x^{n+1}_{\le k}, x^{\leftarrow}_{\le k}), \end{cases} \tag{15}$$

where the second row has to be considered for $0 \le n < N_T$, and the objective function is given by

$$J(v) = \frac{1}{2}\sum_{j=1}^{M} \|Cx^{n(j)} - y_j\|^2,$$

where $t_{n(j)} = \tau_j$, i.e., each measurement time is assumed to correspond to a sampling point of the time grid.

In the direct approach, implicit derivation of scheme (15) with respect to v and applying the Euler scheme to the continuous time equation (10) provides the approximated derivatives x'_k at sampling points t_n by solving

$$\begin{aligned} X_k(m)(x_k'^{n+1} - x_k'^n) = hA_k(v)x_k'^{n+1} + h\partial_v f_k\big(v, x^{n+1}_{\le k}, x^{\leftarrow}_{\le k}\big) \\ + h\sum_{i<k} \partial_{x_i} b_k\big(v, x^{n+1}_{\le i}, x^{\leftarrow}_{\le i}\big)x_i'^{n+1}, \end{aligned} \tag{16}$$

combined with the initial condition $x_k'^0 = 0$ for $k = 1 \dots N$. From the time-dependent state sequence and the time-dependent state derivative sequence, the gradient is obtained by

$$\nabla J(v) = \sum_{j=1}^{M}\sum_{k=1}^{N} (x_k'^{n(j)})^\top C_k^\top (Cx^{n(j)} - y_j). \tag{17}$$

In the adjoint approach, the chosen discrete scheme generates N_T equalities incorporated into the Lagrangian functional by N_T discrete Lagrange multiplier $(\lambda_k^n)_{0\le n<N_T}$. Considering all cumomer fractions of weight k, we obtain the following discrete version of the Lagrangian functional corresponding to (15)

$$\begin{aligned} \mathcal{L}(x,p,\lambda) = \frac{1}{2}\sum_{j=1}^{M}\left\|Cx^{n(j)} - y_j\right\|^2 \\ + \sum_{k=1}^{N}\sum_{n=0}^{N_T-1} (\lambda_k^n)^\top\big(hf_k\big(v, x^{n+1}_{\le k}, x^{\leftarrow}_{\le k}\big) - X_k(m)\big(x_k^{n+1} - x_k^n\big)\big). \end{aligned} \tag{18}$$

It should be emphasized that the discrete version of (12) must be established by deriving the discrete Lagrangian (18), and not simply by choosing discrete schemes corresponding to (12) and (14). The discrete adjoint equation is completely determined once a discrete scheme is chosen for the state equation. For the Euler scheme (15) and the Lagrangian (18), the discrete adjoint scheme is thus given for $n = 1 \dots N_T - 1$ by

$$\begin{aligned} X_k(m)(\lambda_k^n - \lambda_k^{n-1}) = hA_k(v)^\top \lambda_k^{n-1} \\ - h\sum_{i>k} \partial_{x_k} b_i(v, x^n_{<i}, x^{\leftarrow}_{\le i})^\top \lambda_i^{n-1} + \delta_n, \end{aligned} \tag{19}$$

with final condition $\lambda_k^{N_T-1} = 0$ and

$$\delta_n = \begin{cases} C_k^\top(Cx^n - y_j), & \text{if } n = n(j), \\ 0, & \text{otherwise}, \end{cases}$$

which determines $(\lambda_k^n)_{0\le n<N_T}$ for each weight k. Finally, the gradient of J at v is given by

$$\nabla J(v) = h\sum_{k=1}^{N}\sum_{n=0}^{N_T-1} \partial_v f_k\big(v, x^{n+1}_{\le k}, x^{\leftarrow}_{\le k}\big)^\top \lambda_k^n. \tag{20}$$

4.4 Computational Comparison

As in the stationary case, the computation of $\nabla J(v)$ in both the direct and the adjoint approaches involves computing the time-

dependent state x, and the associated cost will depend on the discrete scheme that is chosen. Once again, the difference between the two approaches lies in the cascades (10) and (12): in the direct approach we have N differential equations with p-column matrices, whereas in the direct approach the left-hand side of (12) contains 1-column vectors. Hence, at the very least, we should expect a gain equivalent to the gain obtained in the stationary case.

Another interesting aspect of the adjoint approach appears when the pool sizes are also unknown: in the direct approach equation (10) has to be implicitly derived with respect to m, which leads to a more complex system to express and to solve than (10). In contrast, in the adjoint approach the structure of the system (12) is independent of the choice of parameters. It is only at the final step (14) in the computation of the gradient that the choice of parameters has an impact.

5 Implementation

The methods and algorithms described in this paper have been included in the `sysmetab` software package, distributed under an Open Source license via http://forge.scilab.org/index.php/p/sysmetab/ and available for Linux and MacOSX platforms.

In order to be processed by `sysmetab`, metabolic network description, carbon atom transition map and measurements (stationary or nonstationary Mass Spectrometry (MS) or Nuclear Magnetic Resonance spectroscopy (NMR) data) need to be coded in a plain text XML that respects the FML (Flux Markup Language) format developed for the `13CFlux2` software package [7] (whereas `13CFlux2` only handles stationary problems, the FML input format also allows nonstationary data to be described). For stationary data, `13CFlux2` also supports input files in the older FTBL format, which it automatically converts to FML format. The `sysmetab` software parses the input file and generates a flux identification program in the Scilab language [13] that is specific to the network under consideration. After execution, results are output to a plain text XML file conforming to the Forward Simulation Markup Language developed for the `13CFlux2` software. This file can easily be converted into other file formats, and the fact that `sysmetab` is a command-line tool makes scripting and batch processing straightforward.

Using the adjoint approach is not the only innovative feature of the software. Another novel aspect is the technique chosen for generating the code: starting from the original FML file, the Scilab program is entirely generated using XSL transformations (http://www.w3.org/TR/xslt). These transformations are specified in XSL stylesheets, written in another XML dialect. XSL is very different from other programming languages in use today in that it is a declarative (as opposed to imperative) language. XSL stylesheets have the advantage of being explicit, readable by humans and easy to debug and maintain. The program is generated in several steps, the final step being the transformation of the XML program description into the actual Scilab program. This final step may easily be adapted for other script languages such as Matlab (http:/www.mathworks.com) or Julia (http://julialang.org).

The generated program makes full use of Scilab's potential for improving the speed of calculations: vectorization and sparse matrices are used, and UMFPACK multi-frontal LU factorization [14] is used for calculating the state and adjoint state. The optimization phase is performed using the Feasible Sequential Quadratic Programming (FSQP) algorithm [15], which is available as a Scilab module. Linearized statistics or Monte-Carlo simulations can be used to determine confidence intervals on estimated fluxes. Where Monte-Carlo simulations are used, Scilab allows for parallel execution of the optimization algorithm in a multi-core architecture.

6 Numerical Results

As an example, we considered the central metabolism of *E. Coli*, described in [16] and provided in the network examples in the `influx_s` distribution (the file `e_coli.ftbl`). The experimental stationary data consist of mass spectrometry measurements of the intermediate metabolites Suc, ICit, PEP, PGA, FruBP, Glc6P, Fru6P, Rib5P, Gnt6P, and of the extracellular flux of Acetate. For the adjoint versus direct comparison in the stationary case, the original data were used. For the nonstationary case, synthetic noisy data were generated by a forward nonstationary simulation using the estimated values of flux from the stationary data and some realistic pool sizes (the file `e_coli_ns.fml` in the `sysmetab` distribution).

TABLE 1
`sysmetab` Direct versus Adjoint Average Computation Times and Ratios for the Reduced and Full *E. Coli* Network with Stationary Data

	Direct	Adjoint	Ratio
Reduced network			
State computation (4)	1.5	1.5	1
Cascade (5) versus (7)	6	0.35	17.5
Gradient assembly (6) versus (8)	0.9	0.05	17.2
Total	8.4	1.9	4.47
Full network			
State computation (4)	3.5	3.5	1
Cascade (5) versus (7)	47.2	1.5	31
Gradient assembly (6) versus (8)	4.2	0.2	19
Total	55	5.2	10.6

Time unit = 10^{-3} *s*.

The results presented in this section were obtained on a dedicated Linux server hosting two 10-core Xeon E5-2660 v2 (2.20 GHz) processors with Scilab 5.5.2. Unless explicitly stated otherwise, computations used only one core of the processor, and every reported running time is the median over 10 runs. Although the running times themselves may be very different if different hardware is used, we assume that the ratios between the different running times will remain similar.

6.1 Adjoint versus Direct Gradient Computation

6.1.1 Stationary Data

Table 1 compares the average computation times from steps (5), (6) in the direct approach with steps (7), (8) in the adjoint approach. We report the computation times required for solving the state equation so that the overall times of the two approaches may be compared. In Table 1, time results from the reduced network (using the backwards tracing method [17]) of 1,069 cumomers are contrasted with those obtained from the full original network of 5,455 cumomers.

Since there are almost five times as many cumomers in the full network as in the reduced network, the time needed to compute the state derivative rises. In the direct approach, the contribution of this step to the overall time increases the ratio between the two approaches to approximately 10. In the adjoint approach, the size of the state derivative computation is in relation relation to overall time is smaller. The distribution of overall time reported in Table 1 clearly shows the efficiency of the adjoint computation in comparison with the direct computation.

6.1.2 Nonstationary Data

Using the same network, we obtained estimated flux values from stationary data (given in the next section) and combined them with arbitrary pool size values comprising pool size measurements available in [16]. The resulting values were used to simulate the time course of labeling states. We obtained the data by running the implicit ODE scheme (15) up to a fixed terminal time T, and selecting the simulated measurements at time values $\tau_j, j = 1, \ldots, M$. Gaussian noise was then added by considering the standard deviation values of each of the original stationary labeling state data values.

TABLE 2
`sysmetab` Direct versus Adjoint Average Computation Times and Ratios for the Reduced and Full *E. Coli* Network with Nonstationary Data

	Direct	Adjoint	Ratio
Reduced network			
State computation (15)	21	21	1
Cascade (16) versus (19)	948	15	60
Gradient assembly (17) versus (20)	11	16	0.70
Total	980	53	18.4
Full network			
State computation (15)	87	87	1
Cascade (16) versus (19)	11,214	178	63
Gradient assembly (17) versus (20)	118	90	1.3
Total	11,420	355	32

Time unit = 10^{-3} *s*.

TABLE 3
Estimated Free Flux Values and Final RSS Obtained by `influx_s`, `sysmetab`, and `13CFlux2`

Software	`influx_s`	`sysmetab`	`13CFlux2`
Free fluxes			
Glucupt_1.n	0.80651342	0.80651338	0.80634528
gnd.n	0.14223488	0.14223479	0.13804339
out_Ac.n	0.21300000	0.21300001	0.21299847
pyk.n	1.54425504	1.54425507	1.52116680
zwf.n	0.14233488	0.14233479	0.14145512
ald.x	0.72497526	0.72497650	0.89824047
eno.x	999	999	98,486.1688
fum_a.x	0.43135328	0.43136935	0.40973977
ppc.x	0.17317878	0.17317925	0.17122549
ta.x	0.56653486	0.56653403	0.56756226
tk1.x	0.17878854	0.17878863	0.17125450
tk2.x	0	0	0.00213
Final RSS			
	61.5141593	61.5141593	62.0877059

The different time values comprise 10 equally distributed values up to $\tau_M = 10$ seconds, namely $\tau_j = j$ for $j = 1, \ldots, 10$. We make use of the unconditionally stable property of the implicit Euler scheme in order to set the time step h to $\tau_M/200$. Other experiments show that the outcome x does not vary significantly for smaller values of h.

For the gradient computation, and more generally for the parameter estimations, we chose a value of h larger than that used to generate the simulated data, typically $h = \tau_M/100$, so that the time discretization of the grid is not the same as for the forward problem.

Applying the same computation time analysis to these time-dependent data that we applied in the case of stationary MFA, we report the time distribution of the different steps of the gradient computation: state computation, cascade computation and gradient assembly, namely (15), (16), (17) for the time-dependent direct approach, and (15), (19), (20) for the time-dependent adjoint approach.

For the reduced *E. Coli* network, associated computation times are presented in Table 2. As expected, the time required to compute the state derivative using the direct method (row 2) is greater than the stationary cascade cost (see the second row of Table 1) multiplied by the overall number of time steps (i.e., 100). The cost of the state derivative (direct) cascade in the nonstationary case is in fact more than 150 times the cost of one stationary (direct) cascade. The ratio between the direct approach and the adjoint approach is 60, which demonstrates the efficiency of the adjoint approach. The ratio of over 18 for overall gain is seen to be significantly better than the ratios obtained in the stationary case: the adjoint approach benefits from the nonstationary case, while the direct gradient computation, as expected, is seen to be severely time consuming.

6.2 Assessment of Parameter Estimates

6.2.1 Stationary Data

After comparing the adjoint and the direct approach in `sysmetab` we consider its validation by performing the complete estimation from stationary data. The provided `e_coli.ftbl` file was processed in `influx_s` with the `–emu` command line option. In addition, after conversion to the FML format, it was processed first with the last version of `13CFlux2` with default options and then in `sysmetab` with the command line options `–reg=0` (the default regularization term is not added to the RSS). Table 3 presents the flux values obtained after optimization, (exchange fluxes given by `influx_s` have been converted from normalized $[0, 1[$ values to $[0, +\infty[$ interval)) and the final value of the RSS function $J(v)$.

As the exchange flux eno.x is nearly non-identifiable, the default upper bound is reached by `sysmetab` and `influx_s`. We tried to add the constraint explicitly in the FML file before running `13CFlux2`, but the optimization stopped with a very large value of the RSS. We therefore do not include `13CFlux2` in the discussion below, since we were not able to solve this problem.

As `influx_s` and `sysmetab` give very similar results, we compare their respective behavior in terms of overall computation time, which includes code generation, optimization to obtain the optimal flux values, and linear statistics to obtain their confidence intervals. On the well-defined and quickly converging example of `e_coli.ftbl`, the fastest is `influx_s` (6.7 versus 11.9 seconds for `sysmetab`). But if we only consider optimization time, `sysmetab` is seen to be more efficient (0.35 versus 0.74 seconds for `influx_s`). Comparing sub-second results for a particular network may not be so informative, and we therefore considered a Monte-Carlo estimation of flux statistics with 1,000 data samples, using all available cores of the processors. In this typical situation where the code has to be generated only once, `influx_s` took 105 seconds and `sysmetab` 24 seconds.

We did the same Monte Carlo study involving 1,000 samples on a poorly defined network describing the central metabolism of *E. Coli* and reactions for amino acid biosynthesis (the file `Ecoli.1.ftbl`, available as part of the [8] additional material). Here, `influx_s` shows its superiority in terms of numerical stability. But even though `sysmetab` requires more (but less costly) iterations than `influx_s` for each optimization, it is faster than `influx_s`, terminating in 110 seconds instead of 480.

6.2.2 Nonstationary Data

We are first interested in checking wether the time-dependent labeling measurements are best fitted. Fig. 1 shows the reconstruction of mass isotopomer fractions of measured metabolites using the free fluxes and pool size values obtained after optimization. Comparing these simulated data with the different time measurements marked with circles demonstrates that synthetic data are well recovered. The estimated pool sizes and the estimated flux values obtained from the nonstationary data are reported in Table 4 and contrasted with the fluxes obtained from the stationary data. The 95 percent confidence intervals were computed from the empirical repartition of optimal parameters, estimated by the Monte Carlo method (1,000 resamplings of stationary and nonstationary data were used). We first note that the estimated values obtained from time-dependent labeling state data are similar to those obtained from the stationary state. Then, focusing on the confidence intervals of fluxes, we note that for almost all of them, considering 10 measurements during the transient phase gives better accuracy than using a single measurement when the stationary

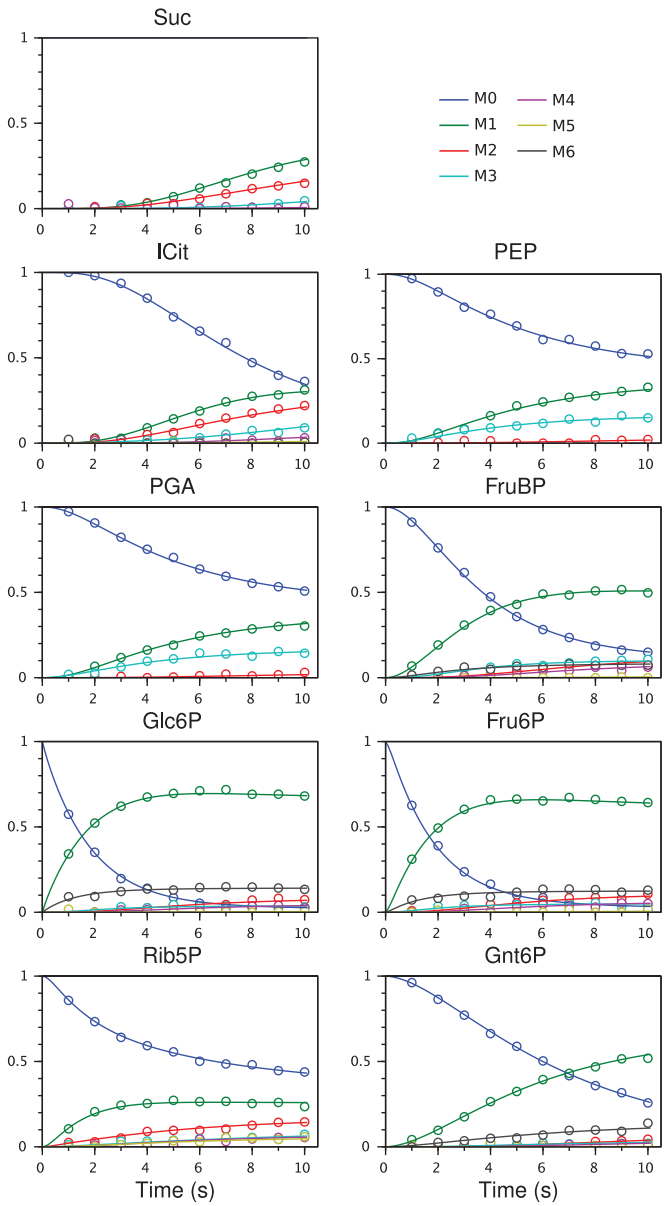

Fig. 1. Experimental mass isotopomers fractions (dots) reconstruction by simulated data (solid lines) after optimization of the parameters.

state is reached. As far as the pool sizes are concerned, for those which have an associated measurement (in bold font), the measurement is well recovered, but apart from Gnt6P all the pools have a significantly larger confidence interval. The choice of measurement times could be further optimized in order to improve the parameter statistics.

If we now compare the computation time on our hardware for the two approaches, we remark that estimating the fluxes and pool size parameters takes almost 8 minutes using the direct method. Using the adjoint method this estimation takes only 36 seconds (21 seconds just for the optimization) and the computation of confidence intervals (Monte Carlo method with 1,000 samples) takes 13 minutes, which demonstrates the efficiency of the approach.

7 Conclusion

In this paper we address the MFA problem by making use of the adjoint state, in the spirit of a general optimization problem governed by a state equation, as in control theory. The proposed methods and algorithms are included in the `sysmetab` software, distributed under an Open Source license. In the stationary case, results demonstrate the efficiency of the approach in terms of precision and computation time when a typical metabolic network is considered. Estimated values were validated against those obtained by other reference software. With the adjoint framework, only vectors (instead of matrices) are updated, which considerably speeds up the computation of the gradient. In the nonstationary case, we considered the same network with synthetic noisy data. On our hardware and using 100 time steps of the ODE integration scheme, a forward simulation followed by the adjoint gradient computation takes 53 milliseconds, instead of almost 1 second when the gradient is computed with the direct method, and a full estimation of fluxes and pool sizes takes 36 seconds. These timings make the nonstationary MFA problem computationally tractable for larger networks and make `sysmetab` competitive with other software.

Although `sysmetab` does not yet implement it, the EMU framework can be used simultaneously with the adjoint approach, and this simultaneous implementation could represent the best computational solution when only MS measurements are used. Although this approach favors gradient-based methods, it can also be used in Gauss-Newton type methods: in this case it allows the computation of the output sensitivity matrix at the cost of n_y adjoint gradients, where n_y is the size of the measurement vector at a given time [12].

Our current development efforts include improving the speed of the code generation step, and implementing high-order ODE solvers and alternative methods for computing nonlinear confidence intervals [18].

TABLE 4
Comparison of Estimated Parameter Values by `sysmetab` from Stationary and Nonstationary Data

Data	Nonstationary		Stationary	
Free fluxes	Median	95% C.I.	Median	95% C.I.
Glucupt_1.n	0.807	[0.803, 0.811]	0.806	[0.798, 0.813]
pyk.n	1.522	[1.293, 1.547]	1.535	[1.355, 1.552]
zwf.n	0.150	[0.131, 0.182]	0.143	[0.114, 0.177]
gnd.n	0.143	[0.128, 0.160]	0.136	[0.101, 0.169]
out_Ac.n	0.213	[0.213, 0.213]	0.213	[0.213, 0.213]
ald.x	0.647	[0.579, 0.717]	0.731	[0.558, 0.929]
eno.x	999	[2.371, 999]	88.61	[0.638, 194.2]
ta.x	0.570	[0.531, 0.611]	0.556	[0.486, 0.635]
tk1.x	0.183	[0.151, 0.214]	0.157	[0.086, 0.222]
tk2.x	0.000	[0.000, 0.021]	0.005	[0.000, 0.066]
fum_a.x	0.100	[0.000, 0.480]	0.310	[0.000, 66.50]
ppc.x	0.241	[0.080, 0.427]	0.189	[0.083, 0.363]
Free pool sizes	Median	95% C.I.		
FruBP	1.860	[1.857, 1.862]		
Glc6P	1.435	[1.428, 1.441]		
Fru6P	0.425	[0.420, 0.431]		
Rib5P	0.020	[0.013, 0.026]		
GA3P	0.470	[0.470, 0.470]		
PEP	0.110	[0.110, 0.110]		
Sed7P	0.057	[0.052, 0.063]		
Suc	0.854	[0.078, 2.427]		
ICit	1.281	[0.376, 1.707]		
PGA	0.478	[0.086, 0.868]		
Gnt6P	0.836	[0.736, 1.022]		
AKG	0.638	[0.000, 1.786]		
Ery4P	0.278	[0.004, 0.619]		
OAA	0.000	[0.000, 1.501]		
Pyr	0.530	[0.000, 2.756]		

Acknowledgments

This work was performed, in partnership with the SAS PIVERT, within the frame of the French Institute for the Energy Transition (Institut pour la Transition Energétique (ITE) P.I.V.E.R.T. (www.institut-pivert.com) selected as an Investment for the Future

("Investissements d'Avenir"). This work was supported, as part of the Investments for the Future, by the French Government under the reference ANR-001.

REFERENCES

[1] W. Wiechert and K. Nöh, "Isotopically non-stationary metabolic flux analysis: Complex yet highly informative," *Current Opinion Biotechnol.*, vol. 24, no. 6, pp. 979–986, 2013.
[2] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, "An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis," *Biotechnol. Bioeng.*, vol. 99, no. 3, pp. 686–699, 2008.
[3] J. D. Young, "INCA: A computational platform for isotopically non-stationary metabolic flux analysis," *Bioinf.*, vol. 30, no. 9, pp. 1333–1335, May 2014.
[4] J. L. Lions and E. Magenes, *Non-Homogeneous Boundary Value Problems and Applications*. New York, NY, USA: Springer-Verlag, 1972.
[5] G. Chavent, "Identification of function parameters in partial differential equations," in *Identification of Parameter Distributed Systems*, R. Goodson and N.-Y. Polis, Eds. New York, NY, USA: ASME, 1974.
[6] R.-E. Plessix, "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications," *Geophys. J. Int.*, vol. 167, pp. 495–503, 2006.
[7] M. Weitzel, K. Nöh, T. Dalman, S. Niedenführ, B. Stute, and W. Wiechert, "13CFLUX2 - high-performance software suite for 13C-metabolic flux analysis," *Bioinf.*, vol. 29, no. 1, pp. 143–145, 2013.
[8] S. Sokol, P. Millard, and J. Portais, "influx_s: Increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments," *Bioinf.*, vol. 28, no. 5, pp. 687–693, 2012.
[9] W. Wiechert and M. Wurzel, "Metabolic isotopomer labeling systems: Part I: Global dynamic behavior," *Math. Biosci.*, vol. 169, no. 2, pp. 173–205, 2001.
[10] A. Cintrón-Arias, H. Banks, A. Capaldi, and A. L. Lloyd, "A sensitivity matrix based methodology for inverse problem formulation," *J. Inverse Ill-Posed Problems*, vol. 17, no. 6, pp. 545–564, 2009.
[11] S. Mottelet. (2012). Fast computation of gradient and sentitivity in 13C metabolic flux analysis instationary experiments using the adjoint method [Online]. Available: http://arxiv.org/abs/1206.5072
[12] A. Sandu and P. Miehe, "Forward, tangent linear, and adjoint RungeKutta methods for stiff chemical kinetic simulations," *Int. J. Comput. Math.*, vol. 87, no. 11, pp. 2458–2479, 2010.
[13] C. Bunks, J. Chancelier, F. Delebecque, C. Gomez, M. Goursat, R. Nikoukhah, and S. Steer, *Engineering and Scientific Computing with Scilab*. Boston, MA, USA: Birkaüser, 1999.
[14] T. A. Davis, "Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method," *ACM Trans. Math. Softw.*, vol. 30, no. 2, pp. 196–199, Jun. 2004.
[15] C. T. Lawrence and A. L. Tits, "A computationally efficient feasible sequential quadratic programming algorithm," *SIAM J. Optimization*, vol. 11, pp. 1092–1118, 2001.
[16] P. Millard, S. Massou, C. Wittmann, J.-C. Portais, and F. Létisse, "Sampling of intracellular metabolites for stationary and non-stationary 13 C metabolic flux analysis in Escherichia coli," *Analytical Biochem.*, vol. 465, pp. 38–49, 2014.
[17] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions," *Metabolic Eng.*, vol. 9, no. 1, pp. 68–86, 2007.
[18] E. M. L. Beale, "Confidence regions in non-linear estimation," *J. Roy. Stat. Soc. Series B (Methodological)*, vol. 22, no. 1, pp. 41–88, 1960.