arXiv:2305.05288v1 [math.OC] 9 May 2023

嵌入全局优化准则的微分代数方程数值模拟

Jens Deussen、Jonathan H¨user 和 Uwe Naumann
信息学12：计算工程软件与工具，亚琛工业大学，德国

文章历史记录
编译于2023年5月10日

摘要W
e是cons id
讨论了嵌入全局优化准则的微分代数方程（DAEO系统），其中嵌入式优化问题通过全局优化求解。对于同时存在多个全局优化器的情况，这实际上会导致微分包含。从一个全局最优解跳转到另一个全局最优解会产生非光滑微分代数方程，从而使数值积分器的收敛阶降至一阶。本文介绍的事件检测与定位方法能够保持积分器较高阶的收敛行为。这使得能够为最优控制问题计算离散切向量和伴随灵敏度。

1. 引言

求解由欠定微分代数方程（DAEs）系统描述的动态模型，也就是说，代数状态多于代数方程的模型，通常需要嵌入优化准则以找到唯一解。这类模型称为嵌入优化准则的微分代数方程（DAEO系统）。

DAEO 的主要应用领域是过程系统工程。具有相间热力学平衡的分离过程通常采用 DAEO 系统进行建模。DAE 描述动态行为，而非线性规划则表示处于吉布斯自由能[1, 8, 29] 最小值时的相平衡。

对于 DAEO 的仿真 [26, 27] 和优化 [4, 16, 28]，通常会用一阶最优性条件，即 Karush-Kuhn-Tucker (KKT) 条件，来替代非线性规划。局部方法只能保证凸非线性规划的精确求解。对于非凸非线性规划，所得解点可能仅为局部最优。此外，用 KKT 条件替代优化问题会导致出现非光滑 DAE，这是由切换事件（即活动集的变化）引起的。此类问题可以采用同时方法或序贯方法 [3] 求解。同时方法需要求解一个单一的大规模非线性规划。相比之下，序贯方法则要求数值积分器与数值优化器相互协作。

联系人：Jens Deussen。电子邮箱：deussen@stce.rwth-aachen.de

本文提出采用确定性全局优化（DGO）方法对 DAEO 系统进行仿真，而不是用一阶最优性条件替代优化问题。这种方法的优点在于，能够保证优化问题的解点是全局最优解。嵌入全局优化问题时，需要将微分包含作为微分方程的一种推广形式，以处理存在多个全局优化器的情况。其中，当全局优化器发生跳变时，也会出现这种情况。全局优化器的跳变被称为事件，这种事件可能由动态系统的固有特性引起。当发生事件时，所得到的 DAE 系统将变得非光滑。即使时间区间涵盖了某个事件，仍使用前一时间步长对应的局部最优点，这就构成了一种不连续性锁定 [25]。遗憾的是，如果不对事件位置进行显式处理，积分器的收敛阶将降低至一阶。为了在全局最优解发生跳变时仍实现二阶收敛，有必要采用自适应时间步进或事件定位过程。本文将介绍如何检测和定位事件。

文献 [13]提出了一种获得积分器更高收敛性的替代方法。该方法利用算法微分（AD）的一种推广形式来处理常微分方程（ODE）的非光滑右端项。有关AD理论的更多信息，请参见 [10, 24] ；有关非光滑AD的信息，请参见 [11, 17]。

本文的组织结构如下：第2节介绍本文所研究问题类型的数学形式化描述，包括对DAEO的描述，以及关于DAEO和事件（即全局极小值点跳变）的假设。第3节概述区间计算和确定性全局优化。基于 [5], ，我们展示如何获得非凸目标函数的所有凸子域，这些子域可能包含局部极小值。第4节描述DAEO的数值模拟，包括如何检测某一时间段内是否发生事件（事件检测），以及如何确定事件发生的时间步长（事件位置）。第5节将所提出的方法应用于两个示例函数。对于第一个示例，我们推导出解析解，以考察在显式处理和不处理事件的情况下仿真的收敛性。第6节总结研究结果，并展望未来工作。

2. 理论背景

我们考虑嵌入全局优化问题的微分包含初值问题。

$$
\begin{aligned}
x(t_0) &= x_0 \\
x'(t) &= f(x(t), y(t)) \\
y(t) &\in \arg\min_{y} h(x(t), y(t)) \quad \forall t \in (t_0, T) \,,
\end{aligned}
\tag{1}
$$

其中 $f\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ 是系统的微分部分， $h\colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}$ 是目标函数。我们用 $x' = \partial x/\partial t$ 表示对时间的导数，用 $\partial_x = \partial/\partial x$ 表示（偏）导数，并用 $\mathrm{d}_x = \mathrm{d}/\mathrm{d}x$表示关于 x 的全导数；在隐式微分的语境中，区分全导数与偏导数至关重要。二阶偏导数记作 $\partial^2_{yx} = \partial^2/\partial y \partial x$。

式 (1) 中的问题是一个微分包含，而不是微分方程问题，因为嵌入其中的优化问题的最优解不一定唯一。本文考虑这样的情形：解集由有限个孤立严格局部最优解构成，且这些解也可能是全局最优解

$$\arg\min_y h(x,y) = \bigcup_{i=1}^{S} \{y^i\} \ ,$$

其中，对于所有 i，$j \in \{1,\dots,S\}$，有 $h(y^i) = h(y^j)$；对于所有 $i \in \{1,\dots,S\}$，$z \in \mathbb{R}^{n_y} \setminus$，有 $h(y^i) < h(z)$ $\arg\min_y h(x,y)$。值得注意的是，这也意味着隐式集值映射$y(t) = y(x(t))$不是凸的。

假设1。函数 f 和 h 关于 (x, y) 二阶利普希茨连续可微。

对于每个严格局部最优解，我们都有必要且充分的局部最优性条件

$$\begin{aligned} 0 &= \partial_y h(x,y) \\ 0 &\prec \partial^2_{yy} h(x,y) \ . \end{aligned}$$

关于 y 的梯度为零，且海森矩阵为正定。

假设2. 初值问题的设定方式应确保对于所有 $i \in \{1,\dots,S\}, t \in [t_0, T]$，$(x(t)\ y^i(t))$ 满足必要且充分的局部最优性条件。

为了将微分包含转化为具有唯一解的不连续微分方程，我们考虑这样一种情形：只有在一个零测集上的时刻，解集的大小才大于 1。为了使这些事件在数值上易于处理，我们进一步作出更强的假设：解集大小大于 1 的时刻构成有限集。

假设3：存在一个有限的事件集合 $0 < t^{e_1} < \cdots < t^{e_K} < T$，其具有

$$\left| \arg\min_y h(x(t^{e_j}), y(t^{e_j})) \right| = 2 \ ,$$

对于所有 $j \in \{1,\dots K\}$，并且我们有

$$\left| \arg\min_y h(x(t), y(t)) \right| = 1 \ ,$$

对于所有 $t \in [t_0, T]$，其中 $t \neq t^{e_j}$ 对某个 $j \in \{1,\dots K\}$ 成立。

根据前述假设，对于每个 t^{e_j}，我们可以分为两种情况。

(i) 从一个全局最优解切换到另一个全局最优解：

$$\arg\min_y h(x(t^{e_j}), y(t^{e_j})) = \{y^1(t^{e_j}), y^2(t^{e_j})\} \ ,$$

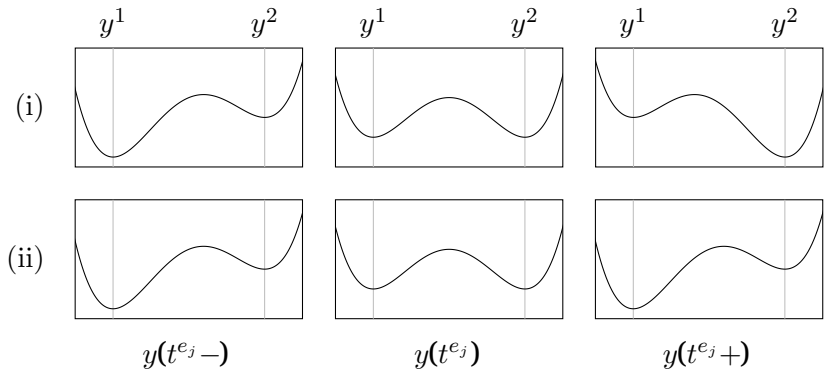

图1.：假设3产生的情形。

and

$$\lim_{t\to t^{e_j}-} \arg\min_y h(x(t), y(t)) = \{y^1(t^{e_j})\} ,$$
$$\lim_{t\to t^{e_j}+} \arg\min_y h(x(t), y(t)) = \{y^2(t^{e_j})\} .$$

可视化结果见图1（上）。

(ii) 与第二个局部最优点相切:

$$\arg\min_y h(x(t^{e_j}), y(t^{e_j})) = \{y^1(t^{e_j}), y^2(t^{e_j})\} ,$$

and

$$\lim_{t\to t^{e_j}-} \arg\min_y h(x(t), y(t)) = \{y^1(t^{e_j})\} ,$$
$$\lim_{t\to t^{e_j}+} \arg\min_y h(x(t), y(t)) = \{y^1(t^{e_j})\} .$$

可视化结果见图1（下）。

第二种情况与微分方程的求解无关，因为它发生在零测集上。我们作出横截性假设，以排除第二种情况。首先，我们将两个局部最优值 y^1 和y^2 接触的条件定义为事件函数 $H(x(t))$ 的根

$$0 = h(x(t), y^1(x(t))) - h(x(t), y^2(x(t))) = H(t) .$$

假设4. 所有事件均为横截的：假设 $y(x(t^{e_j})) = y^1(x(t^{e_j}-))$，则有

$$\partial_x H(x(t^{e_j}))f(x(t), y^1(t^{e_j})) > 0$$
$$\partial_x H(x(t^{e_j}))f(x(t), y^2(t^{e_j})) > 0 \quad j \in \{1, \dots, K\} .$$

在上述假设下，我们可以将原微分包含写成

方程 (1) 中针对 $j = 0, \ldots, K+1$ 的多个微分方程初值问题：

$$\begin{aligned} x^j(t^{e_j}) &= \begin{cases} x_0 & \text{if } j = 0 \\ x^{j-1}(t^{e_j}) & \text{otherwise} \end{cases} \\ (x^j)'(t) &= f(x^j(t), y^j(t)) \\ \{y^j(t)\} &= \arg\min_y h(x^j(t), y^j(t)) \quad t \in (t^{e_j}, t^{e_{j+1}}) \ , \end{aligned} \tag{2}$$

其中 $t^0 = t_0$和 $t^{K+1} = T$。时间段 $(t^{e_j}, t^{e_{j+1}})$、 $j = 0$、.....、 K将称为相。

命题 2.1. 对每个相，方程 (2) 都是一个 DAEO，因为 $y^j(t)$ 是 $x(t)$ 的局部唯一隐函数，且关于 $x(t)$ 利普希茨连续。

证明。我们假设对所有局部最优值都成立的一阶最优性条件为

$$0 = \partial_y h(x, y) \ .$$

根据隐函数定理，并利用海森矩阵的正定性，我们有

$$\begin{aligned} & \mathrm{d}_x \partial_y h(x, y) = \partial^2_{yy} h(x, y) \partial_x y(x) + \partial^2_{yx} h(x, y) = 0 \\ \Rightarrow \quad & \partial_x y(x) = \partial^2_{yy} h(x, y)^{-1} \partial^2_{yx} h(x, y) \ . \end{aligned}$$

□

由于 y 仅以隐式形式给出，上述问题实际上是微分代数方程。通过线性化一阶最优性条件，我们得到微分代数方程的形式：

$$\begin{aligned} x^j(t^{e_j}) &= \begin{cases} x_0 & \text{if } j = 0 \\ x^{j-1}(t^{e_j}) & \text{otherwise} \end{cases} \\ (x^j)'(t) &= f(x^j(t), y^j(t)) \\ 0 &= \partial_y h(x^j(t), y^j(t)) \quad t \in (t^{e_j}, t^{e_{j+1}}) \ . \end{aligned} \tag{3}$$

3. 局部最优值的全局搜索

区间算术（IA） [23] 函数求值具有这样的性质：在给定域上所有可能求得的值，都能可靠地包含在相应区间求值的输出中。为了获得关于函数值的全局信息，在区间算术中只需进行一次函数求值，而不必在多个点进行多次函数求值。

对于（紧）区间变量 $[y]$ ，我们使用以下记号

$$[y] = \left[\underline{y}, \overline{y}\right] = \left\{ y \in \mathbb{R}^{n_y} \mid \underline{y}_i \leq y_i \leq \overline{y}_i \, , \ i = 1, \ldots, n_y \right\} \ ,$$

下界和上界为 $\underline{y}, \overline{y} \in \mathbb{R}^{n_y}$。函数 h 的联合扩展（上确界）

在 $[y]$ 上评估的定义为

$$h^*([y]) = \{h(y) \in \mathbb{R}^{n_g} \mid y \in [y]\} \ .$$

紧致域上代数运算符和初等函数（即一般幂函数、一般根函数、指数函数、对数函数、三角函数和双曲函数）的上确界已广为人知。然而，这并不适用于复合函数。为了对复合函数进行区间分析，区间分析的自然区间扩展（NIE）会在求值过程中用各自的上确界替换所有代数运算符和初等函数。使用 NIE 对 $[y]$ 上的函数 h 求值，得到

$$h([y]) \supseteq h^*([y]) \ .$$

这个超集表明，所得区间可能高估上确界。

如果底层数据格式（例如浮点数）无法表示计算区间的精确边界，就可能出现高估。在这种情况下，区间分析求值会分别将下界和上界向负无穷或正无穷方向舍入。此外，高估也可能由依赖性问题导致。如果函数求值多次使用同一个变量，区间分析不会考虑从这些区间中取出的实际值彼此相等。区间越大，高估就越显著。区间分析面临的另一项挑战是依赖区间自变量的条件分支。对于彼此相交的区间，比较结果并不明确。对原始定义域进行分割或多分割 [15]，并在子域上进行区间分析求值，或许可以解决这一问题。自然区间扩展以线性速度收敛到上确界 [23],，而所谓的中值形式 [22] 则以二次速度收敛。获取函数值全局信息的另一种方法是使用McCormick松弛 [20, 21]。这些松弛以二次速度收敛到凸包，由此可以确定上确界。

为了获得给定域上导数的可靠范围，可以将 NIE 应用于导数计算，例如通过自动微分。在 [7] 中展示了应用于自动微分模型的区间方法的收敛性，并研究了自动微分模型的 NIE 能够给出上确界的情形。我们将区间梯度记为

$$[\partial_y h(x, [y])] \ ,$$

将区间海森矩阵记为

$$\left[\partial^2_{yy} h(x, [y])\right] \ .$$

这些信息可用于排除必要条件不成立的情况，并验证充分条件是否满足。

在 [6], 中，展示了如何使用 McCormick松弛而非 NIE 来收紧梯度包络的界。非光滑的 McCormick松弛是绝对值可分解的，因此我们可以采用 [11]中提出的分段线性化方法。因此，为了计算 McCormick松弛的最优解，我们采用连续分段线性化；这是一种优化方法，通过近端点类型的方法反复使用分段线性化模型 [12]。

A di分而治之算法可用于检测所有凸子域

ˆ$B([y])$，这些子域可能包含 $h(x,y)$ 的局部极小值

$$\hat{B}([y]) = \left\{ [y] \mid 0 \in [\partial_y h\,(x,[y])] \wedge 0 \prec \left[\partial^2_{yy} h\,(x,[y])\right] \right\} .$$

命题 3.2 证明了这些盒子的存在性。

引理 3.1. 对称正定矩阵 A 的谱范数是其最大特征值。

证明。

$$\|A\| = \sqrt{\lambda_{\max}(A^T A)} = \sqrt{\lambda_{\max}(A^2)} = \sqrt{\lambda_{\max}(A)^2} = \lambda_{\max}(A) .$$

□

命题 3.2. 对于每一个局部最优点 $y^\diamond \in \arg\min_y h(x,y)$，都存在一个闭盒

$$B(a,b) = \{y \mid a_i \le y_i - y_i^\diamond \le b_i, i \in \{1,\dots,n_y\}\}$$

其中 $y^\diamond \in \text{int}(B(a,b))$ 满足

$$\partial^2_{yy} h(x,y) \succcurlyeq \delta$$

对于所有 $y \in B\ (a,\ b)$，$\delta > 0$

证明。根据假设1，海森矩阵在该盒域上是利普希茨连续的。利用引理3.1和三角不等式，我们得到

$$\begin{aligned} \left|\lambda_{\max}(\partial^2_{yy} h(x,y)) - \lambda_{\max}(\partial^2_{yy} h(x,y^\diamond))\right| &= \left|\|\partial^2_{yy} h(x,y)\| - \|\partial^2_{yy} h(x,y^\diamond)\|\right| \\ &\le \|\partial^2_{yy} h(x,y) - \partial^2_{yy} h(x,y^\diamond)\| \\ &\le L\|y - y^\diamond\|_2 \end{aligned}$$

对于某些 $L > 0$。根据有限维向量空间中范数的等价性，我们有

$$\|y - y^\diamond\|_2 \le C\|y - y^\diamond\|_\infty \le C \max_{i\in\{1,\dots,n_y\}} b_i - a_i$$

对于某个 $C > 0$。由此可知，对于任意区间盒 $B(a,b)$，我们都能找到一个界限 $Z > 0$，使得

$$\lambda_{\max}(\partial^2_{yy} h(x,y)) \le \lambda_{\max}(\partial^2_{yy} h(x,y^\diamond)) + Y = Z$$

对于所有 $y \in B(a,b)$。令 I 为 $n_y \times n_y$ 单位矩阵。移位矩阵

$$\partial^2_{yy} h(x,y) - ZI$$

对于所有 $y \in B(a,b)$，其具有占优特征值 $\lambda_{\min}(\partial^2_{yy} h(x,y)) - Z$。利用引理 3.1、三角不等式、海森矩阵的强凸性以及利普希茨连续性（As-

假设 1），我们得到

$$\begin{aligned}\lambda_{\min}(\partial^2_{yy}h(x,y)) &= Z - \|\partial^2_{yy}h(x,y) - ZI\| \\ &= Z - \|\partial^2_{yy}h(x,y^\diamond) + (\partial^2_{yy}h(x,y) - \partial^2_{yy}h(x,y^\diamond)) - ZI\| \\ &\geq Z - \|\partial^2_{yy}h(x,y^\diamond) - ZI\| - \|\partial^2_{yy}h(x,y) - \partial^2_{yy}h(x,y^\diamond)\| \\ &\geq Z - \|\partial^2_{yy}h(x,y^\diamond) - ZI\| - L\|y - y^\diamond\| \\ &= \lambda_{\min}(\partial^2_{yy}h(x,y^\diamond)) - L\|y - y^\diamond\| \\ &\geq \mu - L\|y - y^\diamond\| \\ &\geq \mu - LC \max_{i\in\{1,\ldots,n_y\}} b_i - a_i\end{aligned}$$

对所有 $y \in B(a,b)$。我们选择 $b_i > 0$ 和 $a_i < 0$，然后 $y^\diamond \in \text{int}(B(a,b))$，并选择它们使得

$$0 < b_i - a_i \leq \frac{\mu - \delta}{LC}$$

然后，对于所有 $i \in \{1,\ldots,n_y\}$，$\mu - LC(b_i - a_i) \geq \delta$。 □

分治算法递归地细化域，直到某个子域因不满足最优性条件而被消除，或因满足这些条件而被返回。返回的子域与被消除的子域的并集应覆盖原始定义域。处理子域 B 包含以下主要组成部分：

- 消除：若满足以下条件，则丢弃 B

$$0 \notin [\partial_y h\,(x,[y])] \vee 0 \not\prec \left[\partial^2_{yy} h\,(x,[y])\right] .$$

- 终止：若满足以下条件，则将 $\hat{}B([y])$ 添加到 $\mathcal{B}(t)$

$$0 \in [\partial_y h\,(x,[y])] \wedge 0 \prec \left[\partial^2_{yy} h\,(x,[y])\right] .$$

- 分支：将当前域 B 拆分为子域 B_j，其中 $B = \bigcup_j B_j$。

在存储于 $\mathcal{B}(t)$ 中的凸子域上执行局部搜索，得到集合 $\mathcal{S}(t)$，其中包含所有局部极小值 $y^\diamond$。由于前面提到的区间分析（IA）中的高估，某些子域可能不包含局部极小值。在这种情况下，局部搜索会找到子域边界上的极小值，但不会将其纳入集合 $\mathcal{S}(t)$。

为了减少完整搜索算法的计算工作量，可以考虑在算法中加入界定步骤，就像 DGO 求解器（例如 MAiNGO [2] 或 BARON [18]）中实现的传统分支定界算法所做的那样。与其找出所有局部最优值 $y^\diamond$，不如专注于接近全局最优解 y^* 的局部最优值。这些局部最优值将满足

$$h(x,y^\diamond) \leq h(x,y^*) + \alpha\ ,$$

for $\alpha > 0$. 这将导致在处理子域 B时增加额外步骤：

- 界定：计算目标函数在 B 上的凸松弛的保证下界 $\underline{h}$ 。
- 消除：如果满足条件，则丢弃 B。

$$\underline{h} > \overline{h}^* + \alpha \, .$$

- 剪枝：在任意可行点 $y \in B$（例如 B 的中点$y^\dagger$ ）处计算函数值，并在满足 $h(x, y^\dagger) < \overline{h}^*$ 时更新当前最优解 $\overline{h}^*$。

4. 数值模拟

在本节中，我们将介绍如何通过对数值积分器进行时间离散化来数值求解方程 (3)，以及如何检测和定位全局优化器发生切换的事件。

4.1. 时间步进

为了确保动力系统仿真的稳定性，我们在时间步进过程中选择隐式格式；同时，为了提高步长方面的计算效率，采用高阶格式。梯形法则是一种二阶收敛的龙格-库塔方法，它计算的是

$$x(t_{k+1}) = x(t_k) + \frac{\Delta t_k}{2} \left(f\left(x(t_{k+1}), y(t_{k+1})\right) + f\left(x(t_k), y(t_k)\right)\right) \, ,$$

对于离散时间步长 t_k 和 t_{k+1}，其中 $\Delta t_k = t_{k+1} - t_k$ ，从而得到离散化系统

$$\begin{aligned} x^j(t^{e_j}) &= \begin{cases} x_0 & \text{if } j = 0 \\ x^{j-1}(t^{e_j}) & \text{otherwise} \end{cases} \\ x^j(t_{k+1}) &= x^j(t_k) + \frac{\Delta t_k}{2} \left(f\left(x^j(t_{k+1}), y^j(t_{k+1})\right) + f\left(x^j(t_k), y^j(t_k)\right)\right) \\ 0 &= \partial_y h(x^j(t_{k+1}), y^j(t_{k+1})) \quad t_k, t_{k+1} \in (t^{e_j}, t^{e_{j+1}}) \, . \end{aligned} \tag{4}$$

方程（4）中的系统可以通过线性求解器求解，例如使用 LU分解。由于事件 t^{e_j}、$j \in \{1, \ldots, K\}$ 事先未知，因此我们需要一种事件检测机制。

4.2. 事件检测

假设我们跟踪多个强局部最优解的动态集合

$$\mathcal{S}(t) = \{y^1(t), \ldots, y^S(t)\} \, .$$

在 t_k 和 t_{k+1} 之间没有局部最优点出现或消失，因此有 $|\mathcal{S}(t_k)| = |\mathcal{S}(t_{k+1})| = S$。此外，我们假设在 t^e 处存在一个事件。唯一的全局最优解

为 $t_k \le t < t^e$ 是 $y^1(t)$。我们有

$$h(y^1(t)) < h(y^s(t))$$

对于所有 $s \in \{1,\dots,S\} \setminus \{1\}$, $t \in [t_k, t^e)$ 。当在 t^e 处满足无损一般性条件 $h(y^2(t^e)) = h(y^1(t^e))$ 时，便发生一个事件。由于横截性，存在一个 $t_{k+1} > t^e$ ，使得

$$h(y^2(t)) < h(y^s(t))$$

对于所有 $s \in \{1,\dots,S\} \setminus \{2\}$、 $t \in$ （t^e、t_{k+1}]。注意，我们保留了 t_k 中局部最优值的上标，用于 $t \in [t_k, t_{k+1}]$。

离散时间情形下的事件检测问题如下：仅观察通过（全局）优化算法获得的 $\mathcal{S}(t_k)$ 和 $\mathcal{S}(t_{k+1})$ 的随机排序列表表示，在此期间全局最优解是否发生了切换?

我们假设只会发生有限次孤立切换。因此，我们总可以选择足够小的 $\Delta t_k = t_{k+1} - t_k$ ，使每个时间步长内至多发生一次切换。问题不在于 t_k 与 t_{k+1} 之间可能发生了我们无法观察到的来回切换（因为 t_{k+1} 处的最优值又与 t_k 相同）。问题在于如何区分集合 $\mathcal{S}(t_k)$ 和 $\mathcal{S}(t_{k+1})$ 中的元素，从而明确是否发生了一次切换或没有发生切换。

集合 $\mathcal{S}(t_k)$ 和 $\mathcal{S}(t_{k+1})$ 中的元素没有内在顺序。例如，$y^1(t_k)$ 可以是 $\mathcal{S}(t_k)$ 列表表示中的第一个元素，而 $y^1(t_{k+1})$ 则是 $\mathcal{S}(t_{k+1})$ 列表表示中的第二个元素。事件检测的任务归结为区分列表表示中的元素。这样我们就可以检查是否存在某个 $s \neq 1$，使得 $h(y^s(t_{k+1})) < h(y^1(t_{k+1}))$。

令 $\hat{y}_k^1 = y^1(t_k)$ 为 $\mathcal{S}(t_k)$ 的第一个元素。我们希望确定对于 $s \in \{1,\dots,S\}$ ，哪些元素 $\hat{y}^s$ 可能是 $\hat{y}^s = y^1(t_{k+1})$。为此，我们必须考虑到 $y^1(t) = y^1(x(t))$ 是 x 的局部利普希茨连续隐函数。

命题4.1. 给定 $y^1(t_k)$，我们可以用一个利普希茨常数来界定 $y^1(t_k + \Delta t_k)$ 的取值

$$\|y^1(t_k + \Delta t_k) - y^1(t_k)\| \le L\|\Delta t_k\|$$

对于某个 $L > 0$。

证明。由于 Carathéodory 存在性定理 [14], $x(t)$ 是绝对连续的。我们考虑的是一个紧时间区间，因此绝对连续性意味着 Lipschitz 连续性。

$$\|x(t_k + \Delta t_k) - x(t_k)\| \le L_1\|\Delta t_k\| \,.$$

由于强局部凸性，并根据中值定理和隐函数定理，隐函数 $y(x)$ 是利普希茨连续的。

$$y(x + \Delta x) - y(x) = \partial_x y(\xi)\Delta x \le \|\partial_x y(\xi)\|\Delta x$$

对于 some $\xi \in \{x + \lambda\Delta x \mid \lambda \in [0,1]\}$。由于隐函数满足利普希茨连续性

我们得到 $y(x)$

$$\begin{aligned}\|y^1(t_k+\Delta t_k)-y^1(t_k)\| &= \|y^1(x(t_k+\Delta t_k))-y^1(x(t_k))\| \\ &= \|y^1(x(t_k)+x(t_k+\Delta t_k)-x(t_k))-y^1(x(t_k))\| \\ &\le L_2\|x(t_k+\Delta t_k)-x(t_k))\| \\ &\le L_2L_1\|\Delta t_k\| \,.\end{aligned}$$

□

在允许存在一定误差的情况下，同样的思想也适用于离散时间情形。例如，使用显式欧拉步时，这一点显而易见，因为$x(t_{k+1})$与 $x(t_k)$之间的变化由动力学直接给出。我们不使用 t 中的实际利普希茨常数，而是直接使用该时间步长的导数（y^1）′(t_k)。

可通过验证将事件检测功能集成到积分器中。

$$\|y^*(t_k)-y^*(t_{k+1})\| \le \|\Delta t_k\|\cdot\|(y^*)'(t_k)\| \,, \tag{5}$$

在每个时间步长中。如果违反式（5），则检测到一个事件。在这种情况下，接下来需要确定事件位置。

4.3. 事件定位

我们通过与自身假设相兼容的 Mannshardt 方法 [19]，将事件定位过程融入时间步进过程。

求解事件时刻 t^{e_j} 需要通过数值方法求解事件函数的根。我们使用牛顿法求解 $0=H(\tilde{x}(t^{e_j}))$，其中 t^{e_j}，而 $\tilde{x}(t^{e_j})$ 是解轨迹的局部数值近似。对 $\tilde{x}(t^{e_j})$ 关于 t^{e_j}求导很简单，因此我们需要事件函数关于 x的梯度。

$$\begin{aligned}\partial_x H(x) = {} & \partial_x h(x,y^1(x))+\partial_y h(x,y^1(x))\partial_x y^1(x) \\ & -\partial_x h(x,y^2(x))-\partial_y h(x,y^2(x))\partial_x y^2(x) \,.\end{aligned}$$

各个 $\partial_x y^i(x)$均通过上述隐函数定理计算得出。需要注意的是，$y^2(t_{k+1})$可以直接从 DGO 求解器中获得，因为它是全局最优解，而 $y^1(t_{k+1})$则需要在局部最优值集合 $\mathcal{S}($ $t_{k+1})$中进行识别。这可以通过局部优化来实现，即求解

$$\partial_x h(x(t_{k+1},y^1(t_{k+1}))=0 \,,$$

使用初始猜测

$$y^1(t_{k+1})=y^1(t_k)+\Delta t_k\cdot(y^*)'(t_k) \,.$$

5. 数值实验

我们研究了两个全局最优解随时间不连续变化的 DAEO系统。第一个例子非常基础，具有两个局部最优值，且很容易推导出解析解。第二个例子具有多个局部最优值，它们

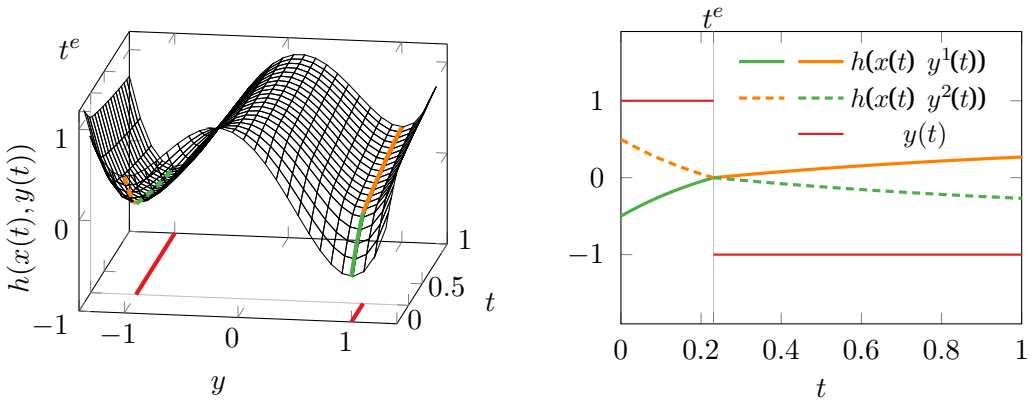

图2：左图显示了 $h(x(t)\ y(t))$ 在 $t \in [0,1]$ 时的曲面图。全局最优解的位置$y(t)$位于 $y^1 = 1$ （ $t < t^e$ 时）和 $y^2 = -1$ （ $t > t^e$ 时），以红线标示，绿线则表示其数值。

会随时间逐渐出现和消失。我们将比较采用和未采用事件检测的数值模拟。此外，我们将数值积分的时间步长设为 $\Delta t = 0.02$ ，并使用上文所述的梯形法则。对于仿真的 DGO 部分，我们使用 [5] 中介绍的求解器。

例5.1。作为第一个例子，我们考虑

$$\begin{aligned} x(0) &= 1 \\ x'(t) &= -(2 + y(t))\, x(t) \\ \{y(t)\} &= \arg\min_y\ h(x(t), y(t)) \ , \end{aligned}$$

其中 $T = 1$ ，且

$$h(x(t), y(t)) = \left(1 - (y(t))^2\right)^2 - \left(x(t) - \frac{1}{2}\right) \sin\left(y(t)\frac{\pi}{2}\right) \ .$$

函数 $h(x(t)\ y(t))$ 在与时间无关的位置 $y^1(t) = 1$ 和 $y^2(t) = -1$ 处有两个局部最优值，即 $\partial_t y^1(t) = 0$ 和 $\partial_t y^2(t) = 0$。图2展示了 $h(x(t)\ y(t))$ 的曲面图。求事件函数的根，即，

$$H(x(t)) = h(x(t), y^1(t)) - h(x(t), y^2(t)) = 0 \ ,$$

在 $x(t) = 0.5$ 处得到事件位置。DAEO 的解为

$$x(t) = \exp(-3t) \ , \ t \in [0, t^e] \ ,$$

第一阶段的

$$\begin{aligned} x(0) &= 1 \\ x'(t) &= -3x(t) \ . \end{aligned}$$

因此，事件位置为 $t^e = -\log(0.5)/3$。第二阶段描述如下：

$$x(t^e) = 0.5$$
$$x'(t) = -x(t)\ ,$$

以及解

$$x(t) = \exp\left(-t + \frac{2}{3}\log(0.5)\right)\ ,\ t \in [t^e, 1]\ .$$

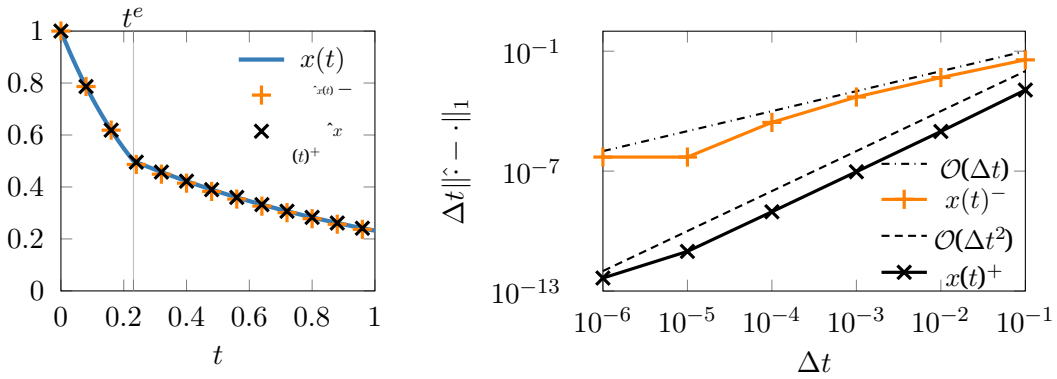

图3：左侧为例5.1中微分变量 $x(t)$ 的解析结果（实线）和数值结果（标记）。微分方程通过梯形法则求解，分别采用不显式处理事件的方式（上标 $-$）和显式处理事件的方式（上标 $+$），其中$\Delta t = 0.02$ 。右侧展示了两种方法的收敛行为。

图3左侧展示了微分变量 $x(t)$ 的结果。蓝线表示解析解，橙色圆圈表示未进行事件检测的数值模拟结果，黑色叉号表示进行了事件检测（及事件位置确定）的数值模拟结果。可以看出，微分变量在事件处并不光滑。图3右侧绘出了随着时间步长减小，显式处理事件（黑色叉号）和不显式处理事件（橙色圆圈）的数值模拟的收敛行为。可以看出，显式处理事件的版本以二次速度收敛至解析解，而未显式处理事件的版本仅以线性速度收敛。只有通过跟踪和检测全局最优解中的切换，才能计算事件位置，并在仿真中显式处理该事件。

例5.2。第二个 DAEO 将一个一维 Griewank 启发函数 [9] 作为嵌入式优化问题。该 DAEO 描述为

$$x(0) = 1$$
$$x'(t) = y(t)$$
$$\{y(t)\} = \arg\min_y\ (y(t) - x(t))^2 + \sin(Cy(t))\ ,$$

与 $C = 5$ 和 $T = 2$。目标函数的多个最优解会随时间出现和消失。左侧给出了该函数及其全局优化器 $y(t)$ 的曲面图

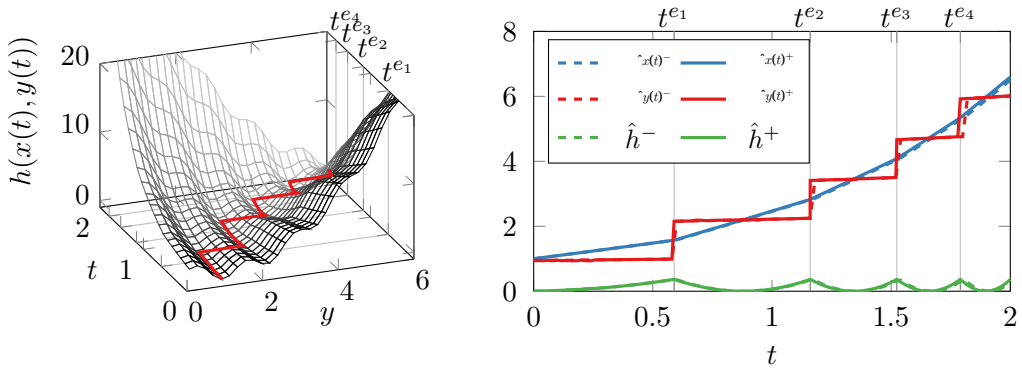

图4：左侧展示了示例例5.2中受 Griewank 启发的一维函数的曲面图及其全局优化器（红线）。右侧展示了采用（实线）和未采用（虚线）事件显式处理时的微分变量 $x(t)$（蓝色）、全局优化器 $y(t)$（红色）以及全局最优解的值（绿色）。

（红线），见图 4 左侧。可以看出，在时间区间 $t \in [0, 2]$ 内有四个事件。可微变量 $x(t)$（蓝色）、全局优化器 $y(t)$（红色）和全局最优解 $h(x(t)\ y(t))$（绿色）在图 4 右侧的图中以曲线表示。未对事件进行显式处理的版本（虚线）所得数值结果与显式处理事件的结果略有差异。这是动力系统行为导致误差传播的结果。

6. 结论与后续工作

我们为跳跃式全局优化器引入了事件检测和事件定位。对事件的显式处理及其在数值积分器中的实现，使得在存在跳跃式全局优化器时，DAEO系统能够采用二阶收敛方法。该二阶积分器可计算动力系统相对于某些参数的离散切线灵敏度和伴随灵敏度，这对求解DAEO系统的最优控制问题至关重要。现在可以通过自动微分方法获得这些灵敏度。

由于在每个时间步长中应用 DGO 都会带来较高的计算成本，因此我们转而致力于随时间跟踪相关的局部最优值。全局搜索仅用于获取局部最优值列表，以及识别在某个时间步长内是否出现了新的局部最优点。采用这种方法，最优解消失不应成为问题。

致谢

本研究得到德国研究联合会（DFG）资助，项目编号为 NA487/8-2。

参考文献

[1] L.E. Baker、A.C. Pierce 和 K.D. Luks，吉布斯能分析与相平衡，SPE J.，22 (1982)，第 731–742 页。[2]D. Bongartz、J. Najman、S. Sass 和 A. Mitsos，MAiNGO——基于 McCormick 的混合整数非线性全局优化算法，技术报告，过程系统工程（AVT.SVT），亚琛工业大学，http://www.avt.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaabclahw，2018年。[3] L.T. Biegler，《非线性规划——概念、算法及其在化学过程中的应用》，工业与应用数学学会，费城，宾夕法尼亚州，2010年。[4] Y. Cao、S. Li、L. Petzold 和 R. Serban，微分代数方程的伴随灵敏度分析：伴随 DAE 系统及其数值解，SIAM J. Sci. Comput. 24 (2003)，第 1076–1089 页。[5] J. Deussen 和 U. Naumann，稀疏高阶导数张量的高效计算，载于《计算科学——ICCS 2019——第19届国际会议》，葡萄牙法鲁，2019年6月12日至14日，第 3–17 页。[6] J. Deussen、J. H¨user 和 U. Naumann，迈向局部最优解的全局搜索，载于《2019年运筹学会议论文集》，J.S. Neufeld、U. Buscher、R. Lasch、D. M¨ost、J. Sch¨onberger 编，施普林格，查姆，2020年，第 97–104 页。[7] J. Deussen，《全局导数》，博士论文，亚琛工业大学，2021年。[8] V. Gopal 和 L.T. Biegler，过程工程中互补问题的平滑方法，AlChE J.，45 (1999)，第 1535–1547 页。[9] A. Griewank，面向全局优化的广义下降法，J. Opt. Th. Appl. 34 (1981)，第 11–39 页。[10] A. Griewank 和 A. Walther，《导数求值：算法微分的原理与技术》，工业与应用数学学会，费城，宾夕法尼亚州，2008年。[11] A. Griewank，论稳定分段线性化与广义算法微分，Optim. Methods Softw.，28 (2013)，第 1139–1178 页。[12] A. Griewank、A. Walther、S. Fiege、S.、T. Bosse，基于灰箱分段线性化的利普希茨优化，Math. Prog.，158 (2016)，第 383–415 页。[13] A. Griewank、R. Hasenfelder、M. Radons、L. Lehmann 和 T. Streubel，利用分段算法微分对利普希茨动力系统进行积分，Optim. Methods Softw.，33 (2018)，第 1089–1107 页。[14] J.K. Hale，《常微分方程》，Robert E. Krieger 出版公司，马拉巴尔，佛罗里达州，1980年。[15] E. Hansen 和 G.W. Walster，《利用区间分析进行全局优化》，Marcel Dekker，纽约，纽约州，2004年。[16] J.L. Hjersted 和 M. A. Henson，利用动态通量平衡模型优化补料分批酿酒酵母发酵，Biotechnol. Progr.，22 (2006)，第 1239–1248 页。[17] J. H¨user，《双曲守恒律不连续解的离散切线与伴随灵敏度分析》，博士论文，亚琛工业大学，2022年。[18] A. Khajavirad 和 N. V. Sahinidis，全局优化松弛的混合 LP/NLP 范式，Math. Prog. Comp.，10 (2018)，第 383–421 页。[19] R. Mannshardt，适用于具有不连续右端项的常微分方程的任意阶单步法，Numer. Math.，31 (1978)，第 131–152 页。[20] G.P. McCormick，可分解非凸规划全局解的可计算性：第一部分——凸下估计问题，Math. Prog.，10 (1976)，第 147–175 页。[21] A. Mitsos、B. Chacuat 和 P.I. Barton，算法的基于 McCormick 的松弛，SIAM J. Opt.，20 (2009)，第 573–601 页。[22] R.E. Moore，《区间分析》，Prentice Hall，恩格尔伍德克利夫，新泽西州，1966年。[23] R.E. Moore、R.B. Kearfott 和 M.J. Cloud，《区间分析导论》，工业与应用数学学会，费城，宾夕法尼亚州，2009年。

[24] U. Naumann, 《计算机程序求导的艺术：算法微分导论》，工业与应用数学学会，费城，宾夕法尼亚州，2012年。[25] T. Park 和 P. I. Barton, 《微分代数模型中的状态事件位置》，《ACM 模型、计算与仿真汇刊》（TOMACS），6（1996年），第137–165页。[26] T. Ploch、E. von Lieres、W. Wiechert、A. Mitsos 和 R. Hannemann-Tam´as, 《Modelica 中嵌入优化准则的微分代数方程组仿真》，《计算机与化学工程》，140（2020年）。[27] T. Ploch、J. Deussen、U. Naumann、A. Mitsos 和 R. Hannemann-Tam´as, 《针对嵌入优化准则的微分代数方程组动态优化的直接单次打靶法》，《计算机与化学工程》，159（2022年）。[28] A.U. Raghunathan、M.S. Diaz 和 L.T. Biegler, 《蒸馏操作动态优化的 MPEC 表述》，《计算机与化学工程》，28（2004年），第2037–2052页。[29] A.M. Sahlodin、H.A. Watson 和 P.I. Barton, 《相变动态仿真的非光滑模型》，《AIChE 杂志》，62（2016年），第3334–3351页。