

高速列车非线性模型的极大似然辨识

衷路生¹ 李兵¹ 龚锦红¹ 张永贤¹ 祝振敏¹

摘要 提出高速列车非线性模型的极大似然 (Maximum likelihood, ML) 辨识方法, 适合于高速列车在非高斯噪声干扰下的非线性模型的参数估计. 首先, 构建了描述高速列车单质点力学行为的随机离散非线性状态空间模型, 并将高速列车参数的极大似然 (ML) 估计问题转化为期望极大 (Expectation maximization, EM) 的优化问题; 然后, 给出高速列车状态估计的粒子滤波器和粒子平滑器的设计方法, 据此构造列车的条件数学期望, 并给出最大化该数学期望的梯度搜索方法, 进而得到列车参数的辨识算法, 分析了算法的收敛速度; 最后, 进行了高速列车阻力系数估计的数值对比实验. 结果表明, 所提出的辨识方法的有效性.

关键词 高速列车, 系统辨识, 极大似然, 平滑滤波器, 梯度搜索

引用格式 衷路生, 李兵, 龚锦红, 张永贤, 祝振敏. 高速列车非线性模型的极大似然辨识. 自动化学报, 2014, 40(12): 2950–2958

DOI 10.3724/SP.J.1004.2014.02950

Maximum Likelihood Identification of Nonlinear Model for High-speed Train

ZHONG Lu-Sheng¹ LI Bing¹ GONG Jin-Hong¹
ZHANG Yong-Xian¹ ZHU Zhen-Min¹

Abstract A maximum likelihood (ML) system identification method is proposed for parameter estimation of nonlinear dynamic high-speed train model subject to non-gaussian noise. Firstly, a stochastic nonlinear discrete state-space model is established to describe the dynamic behavior of high-speed train as a single-point-mass object. The expectation-maximization (EM) approach is employed to compute the ML parameter estimates. In addition, the techniques of particle filtering and particle smoothing are given to estimate the nonlinear state of high-speed train, which is used to compute approximation of the conditional expectation. Furthermore, gradient-based search method is presented to maximize the conditional expectation. And the identification algorithm is given for parameter estimation of high-speed train. The convergence rate of the identification algorithm is also discussed in detail. Finally, numerical simulation study of parameter estimation for high-speed train is implemented and the results show the effectiveness of the proposed ML identification method.

Key words High-speed train, system identification, maximum likelihood (ML), smoothing filters, gradient-based search

Citation Zhong Lu-Sheng, Li Bing, Gong Jin-Hong, Zhang Yong-Xian, Zhu Zhen-Min. Maximum likelihood identification of nonlinear model for high-speed train. *Acta Automatica Sinica*, 2014, 40(12): 2950–2958

收稿日期 2013-10-10 录用日期 2014-03-19
Manuscript received October 10, 2013; accepted March 19, 2014
国家自然科学基金 (61263010, 60904049), 江西省青年科学基金 (20114BA B211014), 江西省教育厅研究项目 (GJJ14399), 国家留学基金 (2011836118) 资助

Supported by National Natural Science Foundation of China (61263010, 60904049), Natural Science Foundation for the Youth of Jiangxi Province (20114BAB211014), Research Project of Jiangxi Educational Department (GJJ14399), and Grant of China Scholarship Council (2011836118)

本文责任编辑 董海荣
Recommended by Associate Editor DONG Hai-Rong
1. 华东交通大学电气与电子工程学院 南昌 330013
1. School of Electrical and Electronic Engineering, East China

高速铁路具有速度高、能耗低、运力大、安全正点等诸多技术经济优势, 因此成为世界各国优先发展的和谐友好绿色交通工具^[1–4]. 然而, 随着列车运行速度的提高, 列车与接触网、轮轨、空气的相互作用显著加剧^[1,4], 列车系统的动态作用环境明显恶化, 这给高速铁路的安全、稳定运行带来巨大隐患. 因此, 为了保障高速铁路的安全运行、促进高速铁路的健康可持续发展, 迫切需要开展适合于高速铁路的有效建模、控制、优化的基础研究.

高速列车的建模、控制、优化问题一直受到中外学者的广泛关注. 其中, 建立高速列车的动力学模型是对列车进行分析、设计、优化的前提和关键. 文献 [2] 建立了包含各节车厢内力作用、暂态影响等耦合因素的多质点–单坐标的高速列车模型, 其建模包括两个步骤: 1) 利用相邻车厢的几何关系把列车的多质点–多坐标模型转换为单坐标模型; 2) 利用“作用力–反作用力”原理得到不含车厢内力的等价模型, 从而消除了由于车厢内力不可测量带来的不便. 基于建立的列车模型, 文献 [2] 进一步提出了计算量适中的鲁棒自适应控制方法, 适合于具有牵引/制动非线性与饱和极限特点的高速列车的速度、位置跟踪问题. 文献 [3] 建立了融合牵引/制动等级、各节车厢内力作用等信息的高速列车多质点模型, 进而设计高速列车的神经自适应容错控制算法, 适用于输入非线性、执行器故障、车厢内力不确定影响等复杂工况. 近年来, 借鉴智能技术能够有效地获取、传递、处理、再生和利用列车运行信息的优势^[4], 相继建立了高速列车的模糊逻辑模型、专家系统模型等智能模型. 我们的前期工作构建了适合于列车子空间预测控制器设计的高速列车子空间预报模型^[4]、时变遗忘因子的增量式子空间预报模型^[5]、在线自适应子空间预报模型^[6]. 列车的子空间预报模型是基于状态空间描述的数据模型, 该模型能够反映高速列车在稳态工作点的牵引/制动特性、专家操作经验等信息, 但列车子空间预报模型本质上是对列车运行过程的局部线性化近似, 难以刻画高速列车的非线性、强耦合等复杂动力学行为.

除了上述面向控制器设计的高速列车建模方法以外, 研究人员还构建了高速列车的多目标优化模型、能耗模型、机械能模型等, 不再赘述.

总而言之, 中外学者从特定需求/目的出发, 对高速列车的建模问题展开了深入研究, 取得了大量适合于特定问题的列车建模的研究成果. 但已有的列车建模方法大多假设列车的模型结构、模型参数、边界条件已知且固定不变, 这难以描述高速列车运行过程的实际动态行为. 例如: 晴朗、雨雪天气下干/湿状态不同的轮轨黏着行为也不相同, 进而影响列车的牵引/制动特性^[6]; 又如: 车速提高后, 车辆、弓网、轮轨之间的动力作用加剧, 列车零部件承受强烈的振动与冲击, 引起列车零部件的材料失效、参数变化和结构破坏, 最终影响列车的动力学性能. 因此, 需要探讨高速列车通用的有效建模方法.

为此, 本文提出高速列车非线性模型的极大似然 (Maximum likelihood, ML) 辨识的建模方法. ML 辨识是基于系统观测数据来建立系统模型的概率统计方法^[7–11], ML 估计满足一致性、渐近性等概率性质, 因此 ML 估计在参数估计、信号处理等领域获得了广泛应用.

具体地, 本文构建了高速列车的单质点模型的离散状态空间系统, 针对列车噪声的非高斯性、列车模型的非线性、列车状态不完全可测等特点, 分析了列车阻力系数 ML 估计的

期望极大 (Expectation maximization, EM) 算法原理, 得到适合于列车状态概率估计的粒子滤波算法和粒子平滑算法, 进而给出列车阻力系数 ML 估计的 EM 实现方法, 并分析了算法收敛速度. 最后进行了高速列车的仿真实验, 结果表明提出的方法是有效的.

1 问题描述

为便于描述, 本文考虑高速列车的单质点模型, 假设线路长度远大于列车的车长, 将列车视为一个刚性质点, 由牛顿力学定律, 得到高速列车在水平直线轨道上的运动方程:

$$\frac{ds}{dt} = v \quad (1a)$$

$$\frac{dv}{dt} = \xi(f(v) - W(v)) \quad (1b)$$

$$\xi = \frac{0.0098}{1+d}; \quad W(v) = a + bv + cv^2$$

其中, s, v 分别是高速列车的位移 (m) 和速度 (m/s); ξ 为列车加速度系数, d 表示列车回转质量系数, 由列车总质量与列车回转部分的折算质量决定, $f(v), W(v)$ 的单位是 N/t (牛顿/吨). 具体地, 可测量的 $f(v)$ 为列车的单位动力; $W(v)$ 为列车受到的单位运行阻力; a 为滚动机械阻力, b 是其他机械阻力系数, c 表示外部空气阻力系数. $W(v)$ 没有统一的计算方法, 通常根据经验公式得到. 不同的高速列车对应的经验系数 (a, b, c) 不相同, 如 CRH3 的 $a = 0.53, b = 0.00392, c = 0.000114$; CRH5 的 $a = 0.693; b = 0.00632; c = 0.000145$.

说明 1.1 式 (1a) 描述的是高速列车的位移、速度关系, 其中的位移可测量; 2) 式 (1b) 描述的是高速列车的力学关系, 其中 $f(v)$ 可测量, $W(v)$ 不可测量, 相对于可测量的位移 s 和单位动力 $f(v)$, $W(v)$ 是缺失数据; 3) $W(v)$ 中的系数 a, b, c 还缺乏有效的计算方法^[2], 并且随着高速列车速度的增加, $W(v)$ 中体现空气阻力的 cv^2 也迅速增大. 另外, 回转质量系数 d 是不固定的 (如乘客数量引起列车总重量变化); 4) 模型 (1a) 和 (1b) 是确定性的模型, 然而列车在实际运行过程中容易受到雨雪、大风、隧道强气流、列车结构件机械/电气特性变化等引起的随机噪声的影响. 因此, 为了科学评估列车的运行性能, 迫切需要探索适合于高速列车参数 $\theta = [a, b, c, d]^T \in \mathbf{R}^4$ 的估计方法, 同时建立描述列车不确定性的随机模型.

式 (1a) 和 (1b) 是列车的连续模型, 但现代检测、监控、通讯等信息技术在高速列车的广泛应用, 使得列车运行过程客观是离散系统. 因此, 需要对列车连续模型 (1a) 和 (1b) 进行离散化. 由于式 (1b) 是非线性的, 因而无法得到模型 (1b) 的解析的离散化方程. 为此, 我们采用前向差分法来近似离散化模型 (1a) 和 (1b), 即

$$\begin{aligned} \frac{ds}{dt} &\approx \frac{s_{k+1} - s_k}{T} \\ \frac{dv}{dt} &\approx \frac{v_{k+1} - v_k}{T} \end{aligned}$$

其中, T 为采样周期.

为了方便后续内容的阐述, 引入控制理论的状态 x_k , 输入 u_k , 输出 y_k 的概念, 即令 $x_k = [s_k, v_k]^T \in \mathbf{R}^2, y_k = s_k \in \mathbf{R}, u_k = f(v_k) \in \mathbf{R}$. 从说明 1 的 1) 和 2) 可知, 状态 x_k 与不可测变量 $W(v)$ 相关, 输入 u_k 和输出 y_k 均可观测.

基于以上分析, 得到对应于列车模型 (1a) 和 (1b) 的随机离散状态空间模型:

$$\begin{aligned} x_{k+1} &= \left(\begin{array}{c} x_{1k} + T x_{2k} \\ x_{2k} + \frac{0.0098}{1+d} \times (u_k - (a + b x_{2k} + c x_{2k}^2)) \end{array} \right) + w_k \\ y_k &= [1 \ 0] x_k + e_k \end{aligned} \quad (2)$$

其中, $x_{1k} = s_k, x_{2k} = v_k$ 分别表示状态 x_k 的第 1 分量和第 2 分量. 显然, 状态的第 1 分量 $x_{1k} = s_k$ 可观测, 而状态的第 2 分量 $x_{2k} = v_k$ 由于含有列车单位运行阻力 $W(v)$ 而不可观测. 因此, 状态 x_k 是不完全可测的. $w_k \in \mathbf{R}^2, e_k \in \mathbf{R}$ 分别是列车的状态和输出受到的随机噪声, 为了描述列车运行环境复杂性 (如雨雪、大风、测量误差等), 噪声 w_k 和 e_k 可以服从非高斯分布, w_k, e_k 的具体分布依据列车的运行工况来确定. 由于引入了随机噪声 w_k 和 e_k , 由模型 (2) 的结构可知其具有马尔科夫的概率性质, 即:

$$x_{k+1} \sim p_{\theta}(x_{k+1}|x_k), \quad y_k \sim p_{\theta}(y_k|x_k) \quad (3)$$

其中, 符号 $p_{\theta}(\Omega|\Delta)$ 表示条件 Δ 下随机向量 Ω 的概率密度. 由列车模型 (2) 可知, $p_{\theta}(x_{k+1}|x_k)$ 等价于状态噪声 w_k 的分布密度 $p_{\theta}(w_k)$, $p_{\theta}(y_k|x_k)$ 等价于输出噪声 e_k 的分布密度 $p_{\theta}(e_k)$.

高速列车的极大似然辨识问题可以描述为: 给定输入/合力-输出/位移的观测序列 D_N ,

$$D_N = \{(u_1, y_1), (u_2, y_2), \dots, (u_N, y_N)\} \quad (4)$$

估计高速列车动态模型 (2) 的参数 $\theta = [a, b, c, d]^T$, 使得式 (5) 定义的输出向量序列 $Y_N = \{y_1, \dots, y_N\}$ 对应的联合概率密度 $p_{\theta}(Y_N)$ 最大

$$p_{\theta}(Y_N) = p_{\theta}(y_1) \prod_{k=2}^N p_{\theta}(y_k|Y_{k-1}) \quad (5)$$

其中, 条件概率密度 $p_{\theta}(y_k|Y_{k-1})$ 表示基于 $k-1$ 时刻的输出/位移向量序列 $Y_{k-1} = \{y_1, \dots, y_{k-1}\}$ 对 k 时刻输出/位移 y_k 的预报概率密度.

2 极大似然估计的 EM 算法原理

为方便计算, 对式 (5) 两边取对数, 把式 (5) 的乘法运算转化为对数求和运算:

$$L(\theta) = \log p_{\theta}(Y_N) = \log p_{\theta}(y_1) + \sum_{k=1}^N \log p_{\theta}(y_k|Y_{k-1}) \quad (6)$$

进一步, 根据概率理论 (全概率为 1) 和列车模型 (2) 的马尔科夫性质, 由贝叶斯定理可推导出式 (5) 和式 (6) 中 $p_{\theta}(y_k|Y_{k-1})$ 的计算方法为

$$p_{\theta}(y_k|Y_{k-1}) = \int p_{\theta}(y_k|x_k) p_{\theta}(x_k|Y_{k-1}) dx_k \quad (7)$$

说明 2. 式 (7) 的优势在于: 利用状态 x_k 关联的两个概率密度 $p_{\theta}(y_k|x_k)$ 和 $p_{\theta}(x_k|Y_{k-1})$ 包含的信息来计算 $p_{\theta}(y_k|Y_{k-1})$.

然而, 由于列车模型 (2) 是非高斯分布的非线性系统, 导致式 (7) 的计算面临三方面困难: 1) 常用的滤波算法均以一定的假设为前提 (如卡尔曼滤波只适用于线性和高斯分布情形), 不适用于估计具有非高斯分布、非线性函数的概率密度 $p_{\theta}(x_k|Y_{k-1})$; 2) 由式 (1b)、(2)、(5) 可知式 (7) 的状态 x_k 中

含有不可测量的单位运行阻力 $W(v_k)$, 即 $W(v_k)$ 是缺失数据, 这也给计算概率密度 $p_{\theta}(y_k|x_k)$ 、 $p_{\theta}(x_k|Y_{k-1})$ 带来了困难; 3) 倘若采用数值计算方法来逼近式 (7) 中关于状态 x_k 的积分, 则运算量大、算法耗时, 难以实时反映高速列车运行的动态行为。

为此, 本文采用期望极大算法来间接优化式 (6) 的对数似然函数, 接下来给出适合于模型 (2) 特点的 EM 算法的实现方案。

EM 算法^[7, 12-13] 将列车模型 (2) 的含有缺失 $W(v_k)$ 的状态序列 $X_N = \{x_1, \dots, x_N\}$ 视为不完全可测数据, 计算由不完全可测数据 X_N 和输出观测数据 Y_N 所组成的全数据 $Z_N = \{X_N, Y_N\} = \{x_1, \dots, x_N, y_1, \dots, y_N\}$ 的联合概率密度, 即:

$$\log p_{\theta}(X_N, Y_N) = \log p_{\theta}(X_N|Y_N)p_{\theta}(Y_N) = \log p_{\theta}(X_N|Y_N) + \log p_{\theta}(Y_N) \quad (8)$$

假设当前由 EM 算法得到的列车参数估计值为 θ_r , 则在 θ_r, Y_N 下对式 (8) 取条件数学期望 $V(\theta, \theta_r)$, 即:

$$V(\theta, \theta_r) = \int \log p_{\theta}(X_N, Y_N) \cdot p_{\theta_r}(X_N|Y_N) dX_N = \int \log p_{\theta}(X_N|Y_N) \cdot p_{\theta_r}(X_N|Y_N) dX_N + \underbrace{(\log p_{\theta}(Y_N)) \cdot \left(\int p_{\theta_r}(X_N|Y_N) dX_N \right)}_{\substack{\text{等于式 (6)} \\ =1}} \quad (9)$$

其中, $p_{\theta_r}(X_N|Y_N)$ 表示列车参数为 θ_r 时以输出序列 Y_N 为条件的状态 X_N 的概率密度。

由式 (9), 进一步可得:

$$\log p_{\theta}(Y_N) - \log p_{\theta_r}(Y_N) = V(\theta, \theta_r) - V(\theta_r, \theta_r) + \underbrace{\left(\int \log \frac{p_{\theta_r}(X_N|Y_N)}{p_{\theta}(X_N|Y_N)} p_{\theta_r}(X_N|Y_N) dX_N \right)}_{term1 \geq 0} \quad (10)$$

式 (10) 中的 $term1 \geq 0$ 是 Kullback-Leibler 信息测度的结论^[7, 13]。

根据式 (10), 可知采用 EM 算法进行高速列车模型 (2) 的极大似然估计的关键: 选择参数 θ 使得 $V(\theta, \theta_r) - V(\theta_r, \theta_r) > 0$, 就能保证 $\log p_{\theta}(Y_N) - \log p_{\theta_r}(Y_N) > 0$ (即模型 (2) 的联合概率密度是增加的)。

基于以上分析, 得到利用 EM 算法获得模型 (2) 的极大似然辨识的迭代计算步骤为:

步骤 1. 构造条件数学期望 $V(\theta, \theta_r)$;

步骤 2. 最大化条件数学期望, 即求解优化问题:

$$\theta_{r+1} = \arg \max_{\theta} V(\theta, \theta_r)$$

3 列车状态条件概率密度的估计

为了利用列车模型 (2) 状态方程、输出方程所包含的信息, 将式 (9) 定义的 $V(\theta, \theta_r)$ 等价表示为

$$V(\theta, \theta_r) = \int \log p_{\theta}(X_N) \cdot p_{\theta_r}(X_N|Y_N) dX_N + \int \log p_{\theta}(Y_N|X_N) \cdot p_{\theta_r}(X_N|Y_N) dX_N = I_1 + I_2 \quad (11)$$

其中, I_1, I_2 分别为

$$I_1 = \sum_{k=1}^N \int \log p_{\theta}(y_k|x_k) p_{\theta_r}(x_k|Y_N) dx_k \quad (12a)$$

$$I_2 = \sum_{k=1}^{N-1} \int \int \log p_{\theta}(x_{k+1}|x_k) p_{\theta_r}(x_{k+1}, x_k|Y_N) dx_k dx_{k+1} \quad (12b)$$

由式 (9)、(11)、(12) 可知, 计算 I_1 是与条件分布密度 $p_{\theta_r}(x_k|Y_N)$ 有关的一重积分, 计算 I_2 是与条件分布密度 $p_{\theta_r}(x_{k+1}, x_k|Y_N)$ 有关的二重积分。从统计学角度, 获取 $p_{\theta_r}(x_k|Y_N)$ 、 $p_{\theta_r}(x_{k+1}, x_k|Y_N)$ 的信息属于多元密度估计的范畴。常用的多元密度估计方法是多变量核密度估计器, 该方法需要对每个变量设计合适的带宽。然而, 计算带宽通常是非线性、非凸的优化问题, 难以得到最优的带宽参数。

为此, 本文借鉴粒子滤波、粒子平滑的思想估计列车的条件分布密度。1) 采用粒子滤波器估计条件分布密度 $p_{\theta}(x_k|Y_k)$; 2) 采用粒子平滑器估计 $p_{\theta}(x_k|Y_N)$ 、 $p_{\theta_r}(x_{k+1}, x_k|Y_N)$ 。

3.1 列车的粒子滤波器

粒子滤波是基于蒙特卡洛、贝叶斯推理和序贯重要性再采样的滤波方法^[11, 13], 其基本思想是利用重要性密度 $q(\cdot)$ 中 M 个独立的采样/粒子 $\{\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^M\}$ 及对应的滤波权重 $\{\beta^1, \dots, \beta^M\}$ ($\sum_{i=1}^M \beta^i = 1$) 来非参数化地逼近目标概率密度 $\pi(\cdot)$ 。

针对列车模型 (2), 同时考虑到式 (9)、(11)、(12), 选取目标概率密度 $\pi(\cdot)$ 、重要性密度 $q(\cdot)$ 分别为

$$\pi(x_k) = p_{\theta}(x_k|Y_k) \quad (13a)$$

$$q(\tilde{x}_k^i) = p_{\theta}(\tilde{x}_k^i|x_{k-1}) \quad (13b)$$

$$\pi(x_k) \approx \sum_{i=1}^M \beta_k^i q(\tilde{x}_k^i), \quad \sum_{i=1}^M \beta_k^i = 1 \quad (13c)$$

说明 3. 式 (13b) 的重要性密度具有可行性, 因为其实现 $\tilde{x}_k^i \sim p_{\theta}(\tilde{x}_k^i|x_{k-1})$ 可以通过先验噪声分布 $\tilde{w}_k^i \sim p_{\theta}(\tilde{x}_k^i|x_{k-1})$ 来完成, 其中 \tilde{w}_k^i 为模型 (2) 中对应于状态 \tilde{x}_k^i 的噪声。

接下来讨论 (13c) 的滤波权重 $\beta_k^i, i = 1, \dots, M$ 的计算。

根据贝叶斯定理、列车模型 (2) 的马尔科夫性质以及概率理论,

$$\frac{\pi(\tilde{x}_k^i)}{q(\tilde{x}_k^i)} = \frac{p_{\theta}(\tilde{x}_k^i|Y_k)}{\int p_{\theta}(\tilde{x}_k^i|x_{k-1}) p_{\theta}(x_{k-1}|Y_{k-1}) dx_{k-1}} = \quad (14a)$$

$$\frac{p_{\theta}(y_k|\tilde{x}_k^i)}{p_{\theta}(y_k|Y_{k-1})} \quad (14b)$$

式 (14a) 用到了全概率的性质。式 (14b) 的分子 $p_{\theta}(y_k|\tilde{x}_k^i)$ 可以根据列车模型 (2) 的输出/位移方程 $y_k = s_k + e_k$ 得到, 式 (14b) 的分母 $p_{\theta}(y_k|Y_{k-1})$ 与粒子 \tilde{x}_k^i 无关。

基于以上分析, 得到式 (13c) 的滤波权重的一种计算方法:

$$\beta_k^i = \frac{1}{\lambda} p_{\theta}(y_k|x_k^i), \quad \lambda = \sum_{i=1}^M \beta_k^i \quad (15)$$

至此, 完成了对于列车模型 (2) 的粒子滤波器的设计。然而, 由于粒子滤波算法面临着“粒子退化”的共性问题^[14], 因而, 需要选择新的粒子 \tilde{x}_k^j 进行再采样, 再采样使得滤波权重高的粒子保留下来, 同时删除权重低的粒子。

综合上述分析, 得到适合于列车模型 (2) 的粒子滤波算法.

算法 1. 粒子滤波算法

- 步骤 1. 初始化粒子 $\{x_0^i\}_{i=1}^M \sim p_{\theta}(x_0)$, 置 $k = 1$;
 步骤 2. 根据列车模型 (2) 的状态方程更新粒子:

$$\tilde{x}_{k+1}^i \sim p_{\theta}(x_{k+1}|\tilde{x}_k^i) \quad (16)$$

步骤 3. 依据式 (15), 计算粒子滤波权重序列 β_k^i , $i = 1, \dots, M$;

步骤 4. 选择新的粒子进行再采样;

步骤 5. 如果 $k < N$, 则 $k = k + 1$, 返回步骤 2, 否则, 算法结束.

通过以上步骤 1~5, 得到了目标分布密度 $\pi(x_k) = p_{\theta}(x_k|Y_k)$ (即以列车输出/位移序列 $Y_k = \{y_1, \dots, y_k\}$ 为条件的列车状态 x_k 的分布密度) 的近似逼近:

$$p_{\theta}(x_k|Y_k) \approx \hat{p}_{\theta}(x_k|Y_k) = \sum_{i=1}^M \beta_k^i \delta(x_k - \tilde{x}_k^i) \quad (17)$$

其中, 如果 $x_k \geq \tilde{x}_k^i$, 则 $\delta(x_k - \tilde{x}_k^i) = 1$; 否则 $\delta(x_k - \tilde{x}_k^i) = 0$.

说明 4. 以上步骤 2 和步骤 4 均与选择的随机数的概率密度有关, 即步骤 2 和步骤 4 得到的序列是不确定的, 这将影响粒子平滑器的计算时间 (具体分析见第 3.2 节的说明 5).

一旦获得列车状态的条件密度 $p_{\theta}(x_k|Y_k)$ 的估计, 就完成了粒子滤波器的设计, 据此可以进一步设计列车的粒子平滑器.

3.2 列车的粒子平滑器

为了计算式 (9)、(11)、(12), 尚需要进一步估计以列车所有输出/位移序列 $Y_N = \{y_1, \dots, y_N\}$ 为条件的列车状态 x_k 的平滑分布密度 $p_{\theta}(x_k|Y_N)$, 这一问题通常被称为粒子平滑问题.

为此, 先对列车模型 (2) 的 $p_{\theta}(x_k|Y_N)$ 进行分析, 根据全概率性质、列车模型 (2) 的马尔科夫性质以及贝叶斯定理可得:

$$p_{\theta}(x_k|Y_N) = \int p_{\theta}(x_k|x_{k+1})p_{\theta}(x_{k+1}|Y_N)dx_{k+1} = \int p_{\theta}(x_k|x_{k+1}, Y_k)p_{\theta}(x_{k+1}|Y_N)dx_{k+1} \quad (18a)$$

$$p_{\theta}(x_k|Y_k) \underbrace{\left(\int \frac{p_{\theta}(x_{k+1}|x_k)p_{\theta}(x_{k+1}|Y_N)}{p_{\theta}(x_{k+1}|Y_k)} dx_{k+1} \right)}_{term2} \quad (18b)$$

式 (18b) 中的列车滤波密度 $p_{\theta}(x_k|Y_k)$ 可根据式 (17) 计算, 接下来进一步计算式 (18b) 中的积分项. 由全概率性质, 式 (18b) 中 term2 的分母 $p_{\theta}(x_{k+1}|Y_k)$ 可用积分表示为

$$p_{\theta}(x_{k+1}|Y_k) = \int p_{\theta}(x_{k+1}|x_k)p_{\theta}(x_k|Y_k)dx_k \quad (19)$$

由于列车状态 x_k 包含不可测的列车阻力, 无法直接计算式 (19) 关于状态 x_k 的积分. 为此, 借鉴第 3.1 节概率密度逼近的思想, 引入第 3.1 节的重要性密度 $q(\cdot)$ 的粒子 $\{\tilde{x}^1, \dots, \tilde{x}^M\}$ 来对式 (19) 的积分区间进行分割, 从而将式 (19) 关于 x_k 的积分问题转化为关于粒子 $\{\tilde{x}^1, \dots, \tilde{x}^M\}$ 的数值求和问题, 结合式 (17), 有

$$p_{\theta}(x_{k+1}|Y_k) = \sum_{i=1}^M p_{\theta}(x_{k+1}|\tilde{x}_k^i)p_{\theta}(\tilde{x}_k^i|Y_k) = \sum_{i=1}^M \beta_k^i p_{\theta}(x_{k+1}|\tilde{x}_k^i) \quad (20)$$

其中的权重 β_k^i 已经在第 3.1 节的粒子滤波算法的步骤 3 计算得到, $p_{\theta}(x_{k+1}|\tilde{x}_k^i)$ 可由列车模型 (2) 的状态方程得到.

类似地, 引入新的粒子对式 (18b) 关于状态 x_{k+1} 的积分区间进行分割, 得到式 (18b) 的数值求和的关系式为

$$p_{\theta}(x_k|Y_N) \approx \hat{p}_{\theta}(x_k|Y_N) = \sum_{i=1}^M \beta_{k|N}^i \delta(x_k - \tilde{x}_k^i) \quad (21)$$

$$\beta_{k|N}^i = \beta_k^i \sum_{j=1}^M \beta_{k+1|N}^j \frac{p_{\theta}(\tilde{x}_{k+1}^j|\tilde{x}_k^i)}{\sum_{i=1}^M \beta_k^i p_{\theta}(\tilde{x}_{k+1}^j|\tilde{x}_k^i)}$$

其中, $\beta_{k|N}^i$ 表示以列车所有输出/位移序列 $Y_N = \{y_1, \dots, y_N\}$ 为条件的粒子 \tilde{x}_k^i 的平滑权重.

综合以上分析, 得到适合于列车模型 (2) 的粒子平滑算法.

算法 2. 粒子平滑算法

步骤 1. 执行列车的粒子滤波算法 1, 保存粒子 $\{\tilde{x}_k^i\}_{i=1}^M$ 及其对应的滤波权重 $\{\beta_{k+1}^i\}_{i=1}^M$, $k = 1, \dots, N$;

步骤 2. 利用 $k = N$ 对应的粒子滤波权重 β_N^i 来初始化列车的粒子平滑权重 $\beta_{N|N}^i$, 即令 $\beta_{N|N}^i = \beta_N^i$, $i = 1, \dots, M$, 并置 $k = N - 1$;

步骤 3. 根据式 (21) 计算列车的粒子平滑权重序列 $\{\beta_{k|N}^i\}_{i=1}^M$;

步骤 4. 如果 $k > 0$, 则 $k = k - 1$, 返回步骤 3, 否则, 算法结束.

说明 5. 步骤 3 本质上是基于 k 时刻的信息来搜索满足概率要求的 $(k - 1)$ 时刻的粒子, 并且步骤 1 的粒子滤波具有随机性 (详见说明 4), 这增加了步骤 3 的搜索时间. 因此, 本文给步骤 3 设置搜索时间的阈值 $Tmax$, 如果步骤 3 的搜索时间大于 $Tmax$, 则返回步骤 1 (具体实现见第 4.3 节的列车辨识算法).

4 列车参数估计的 EM 算法实现

列车模型 (2) 的参数 $\theta = [a, b, c, d]^T$ 的极大似然估计的 EM 算法有两个关键步骤: 1) 构造列车的条件数学期望; 2) 最大化列车的条件数学期望.

4.1 构造列车的条件数学期望

由第 2 节和第 3 节可知, 构造列车的条件数学期望, 即计算式 (11) 的 I_1 和 I_2 . 从式 (12) 可知, I_1 和 I_2 的特点为: I_1 是以列车状态 x_k 为积分变量的一重积分, 并且 I_1 与条件平滑密度 $p_{\theta_r}(x_k|Y_N)$ 有关 (其中 θ_r 是利用 EM 算法得到的列车当前参数的估计); I_2 是以列车状态 x_k 和 x_{k+1} 为积分变量的二重积分, 且 I_2 与联合条件平滑密度 $p_{\theta_r}(x_k, x_{k+1}|Y_N)$ 有关. 类似于第 3 节的分析, 由于列车状态 x_k 和 x_{k+1} 均含有不可测的运行阻力, 难以获得 I_1 和 I_2 的解析计算方法.

为此, 借鉴第 3 节的思想, 我们融合粒子滤波和粒子平滑^[11]的方法来计算式 (11) 的 I_1 和 I_2 .

式 (11) 和式 (12) 中的 I_1 是以列车状态 x_k 为积分量的一重积分, 并且 I_1 中的条件平滑密度 $p_{\theta_r}(x_k|Y_N)$ 可由式 (21) 计算. 因此, 借鉴第 3 节的思路, 得到基于粒子平滑器的

I_1 的数值计算方法为

$$I_1 = \sum_{k=1}^N \int \log p_{\theta}(y_k|x_k) p_{\theta_r}(x_k|Y_N) dx_k \approx \sum_{k=1}^N \sum_{i=1}^M \beta_{k|N}^i \log p_{\theta}(y_k|\tilde{x}_k^i) = \hat{I}_1 \quad (22)$$

其中, $\beta_{k|N}^i$ 根据算法 1 和算法 2 计算.

式 (11) 和式 (12) 中的 I_2 涉及到两个连续时刻的状态变量 x_k, x_{k+1} 的联合条件平滑密度 $p_{\theta_r}(x_{k+1}, x_k|Y_N)$, 而第 3 节的结果仅适用于估计单变量的条件分布密度. 因此, 首先对联合条件平滑密度 $p_{\theta_r}(x_{k+1}, x_k|Y_N)$ 进行等价变换, 使之能套用第 3 节的方法来近似计算 I_2 . 具体地, 根据贝叶斯定理有

$$p_{\theta_r}(x_{k+1}, x_k|Y_N) = \frac{p_{\theta_r}(x_{k+1}, x_k, Y_N) p_{\theta_r}(x_{k+1}, Y_N)}{p_{\theta_r}(x_{k+1}, Y_N) p_{\theta_r}(Y_N)} = \underbrace{(p_{\theta_r}(x_k|x_{k+1}, Y_N))}_{term3} p_{\theta_r}(x_{k+1}|Y_N) \quad (23)$$

进一步, 把列车所有输出/位移序列 Y_N 分解为 $Y_N = \{Y_k, Y_{k+1:N}\}$, 将式 (23) 的 $term3$ 等价

$$term3 = p_{\theta_r}(x_k|x_{k+1}, Y_k, Y_{k+1:N}) = \frac{p_{\theta_r}(Y_{k+1:N}|x_k, x_{k+1}, Y_k) p_{\theta_r}(x_k|x_{k+1}, Y_k)}{p_{\theta_r}(Y_{k+1:N}|x_{k+1}, Y_k)} = \quad (24a)$$

$$p_{\theta_r}(x_k|x_{k+1}, Y_k) = \quad (24b)$$

$$\frac{p_{\theta_r}(x_{k+1}|x_k) p_{\theta_r}(x_k|Y_k)}{p_{\theta_r}(x_k|Y_k)} \quad (24c)$$

式 (24a) 到式 (24b) 利用了列车模型 (2) 的马尔科夫性质, 即 $k+1$ 时刻的状态 x_{k+1} 包含了 k 时刻的状态 x_k 的信息, 因而式 (24a) 分子和分母的条件概率密度满足等式关系 $p_{\theta_r}(Y_{k+1:N}|x_{k+1}, Y_k) = p_{\theta_r}(Y_{k+1:N}|x_k, x_{k+1}, Y_k)$, 两者抵消后即得到式 (24b).

把式 (24c) 代入式 (23) 可得:

$$p_{\theta_r}(x_k, x_{k+1}|Y_N) = \frac{p_{\theta_r}(x_{k+1}|x_k) p_{\theta_r}(x_k|Y_k) p_{\theta_r}(x_{k+1}|Y_N)}{p_{\theta_r}(x_{k+1}|Y_k)} \quad (25)$$

至此, 已经将联合条件概率密度 $p_{\theta_r}(x_k, x_{k+1}|Y_N)$ 等价变换为式 (25) 等号右边的单变量条件概率密度的乘积, 从而可以套用第 3 节的方法来近似计算式 (11) 的二重积分 I_2 . 为此, 选用两组粒子 (标号分别为 i, j) 对 I_2 的积分变量 x_k, x_{k+1} 进行离散, 同时将式 (25) 代入 I_2 得到:

$$I_2 = \sum_{k=1}^{N-1} \iint \log p_{\theta}(x_{k+1}|x_k) p_{\theta_r}(x_k, x_{k+1}|Y_N) dx_k dx_{k+1} \approx \sum_{k=1}^{N-1} \sum_{i=1}^M \beta_k^i \int \frac{p_{\theta_r}(x_{k+1}|Y_N) p_{\theta_r}(x_{k+1}|\tilde{x}_k^i)}{p_{\theta_r}(x_{k+1}|Y_k)} \times \log p_{\theta}(x_{k+1}|\tilde{x}_k^i) dx_{k+1} = \sum_{k=1}^{N-1} \sum_{i=1}^M \sum_{j=1}^M \beta_k^i \beta_{k|N}^j \frac{p_{\theta_r}(x_{k+1}^j|\tilde{x}_k^i) \log p_{\theta}(x_{k+1}^j|\tilde{x}_k^i)}{p_{\theta_r}(\tilde{x}_{k+1}^j|Y_k)} = \hat{I}_2 \quad (26)$$

式中的分母 $p_{\theta_r}(\tilde{x}_{k+1}^j|Y_k)$ 可以根据式 (20) 来计算.

至此, 已经完成了对于列车联合条件数学期望 (11) 的近似计算, 即

$$V(\theta, \theta_r) \approx \hat{V}(\theta, \theta_r) = \hat{I}_1 + \hat{I}_2 \quad (27)$$

4.2 最大化列车的条件数学期望

由于列车模型 (2) 是非线性函数, 且列车噪声可以服从非高斯分布, 导致第 4.1 节构造的列车的联合条件数学期望 $\hat{V}(\theta, \theta_r)$ 与列车参数 θ 是非线性、非凸关系. 对于此类问题通常用智能优化方法^[10] (如耦合模拟退火算法 (Coupled simulated annealing, CSA)、粒子群算法 (Particle swarm optimization, PSO)、遗传算法等) 进行求解. 智能优化通常具有全局收敛、算法稳定等优点, 但也存在算法速度慢、算法关键参数 (如 CSA 算法的接受温度、生成温度) 的设置缺乏理论依据等不足.

为此, 借鉴课题组的前期工作^[3, 5, 8-9], 本文提出梯度搜索的方法来优化第 4.1 节构造列车联合条件数学期望 $\hat{V}(\theta, \theta_r)$.

将式 (27) 在列车当前参数 θ_r 进行二阶泰勒级数展开, 得到

$$\hat{V}(\theta, \theta_r) = \hat{V}(\theta_r, \theta_r) + \nabla \hat{V}|_{\theta_r}(\theta - \theta_r) + \frac{1}{2}(\theta - \theta_r)^T \nabla^2 \hat{V}|_{\theta_r}(\theta - \theta_r) + o(\cdot) \quad (28)$$

其中, $o(\cdot)$ 表示高阶小量, $\nabla \hat{V}|_{\theta_r} \in \mathbf{R}^4$ 是 $\hat{V}(\theta, \theta_r)$ 对于列车参数 $\theta = [a, b, c, d]^T$ 的一阶导数向量, $\nabla^2 \hat{V}|_{\theta_r} \in \mathbf{R}^{4 \times 4}$ 表示 $\hat{V}(\theta, \theta_r)$ 对于列车参数 θ 的二阶导数矩阵, $\nabla \hat{V}|_{\theta_r}$ 和 $\nabla^2 \hat{V}|_{\theta_r}$ 的定义分别为

$$\nabla \hat{V}|_{\theta_r} = \frac{\partial \hat{V}(\theta, \theta_r)}{\partial \theta} |_{\theta_r} \quad (29)$$

$$\nabla^2 \hat{V}|_{\theta_r} = \frac{\partial^2 \hat{V}(\theta, \theta_r)}{\partial \theta \partial \theta^T} |_{\theta_r} \quad (30)$$

由于 $\hat{V}(\theta, \theta_r)$ 是通过粒子平滑算法和粒子滤波算法构造得来, 并且 $\hat{V}(\theta, \theta_r)$ 与列车参数 θ 是非线性、非凸关系, 因而难以得到 $\nabla \hat{V}|_{\theta_r}$ 和 $\nabla^2 \hat{V}|_{\theta_r}$ 的解析表达式. 为此, 采用数值计算^[8] 的方法来逼近 $\nabla \hat{V}|_{\theta_r}$ 和 $\nabla^2 \hat{V}|_{\theta_r}$. 以 $\nabla \hat{V}|_{\theta_r}$ 为例:

$$\nabla \hat{V}(i)|_{\theta_r} \approx \lim_{\epsilon \rightarrow 0} \frac{\hat{V}(\theta_r + \epsilon \theta_{r,i}, \theta_r) - \hat{V}(\theta_r, \theta_r)}{\epsilon} \quad (31)$$

式中的 $\theta_{r,i}$ ($i = 1, 2, 3, 4$) 表示向量 θ 的第 i 个分量. 根据导数的极限定义, 式 (31) 中的 ϵ 越小, 则其对导数的逼近精度越高.

同理, 可以写出 $\nabla^2 \hat{V}|_{\theta_r}$ 的数值计算式.

在计算出 $\nabla \hat{V}|_{\theta_r}$ 和 $\nabla^2 \hat{V}|_{\theta_r}$ 以后, 如果忽略式 (28) 的高阶小量 $o(\cdot)$, 则式 (28) 是关于列车参数向量 θ 的函数. 因而 $\hat{V}(\theta, \theta_r)$ 的最大化问题就转化为式 (28) 的关于向量 θ 的极值问题. 为此, 将式 (28) 对 θ 求偏导数, 并令偏导数为零向量, 即:

$$\frac{\partial \hat{V}(\theta, \theta_r)}{\partial \theta} = \nabla \hat{V}|_{\theta_r} + \nabla^2 \hat{V}|_{\theta_r}(\theta - \theta_r) = \mathbf{0}_{4 \times 1} \quad (32)$$

由式 (32), 可以获得列车参数向量 θ 的解, 把该解赋给 θ_{r+1} , 则得到第 $r+1$ 步列车参数向量的表达式为

$$\theta_{r+1} = \theta = \theta_r - (\nabla^2 \hat{V}|_{\theta_r})^{-1} \nabla \hat{V}|_{\theta_r} \quad (33)$$

式 (33) 给出了最大化列车条件数学期望的参数向量的更新方法.

4.3 列车参数的辨识算法

综合前面各节的分析, 得到列车模型 (2) 参数的极大似然辨识算法.

算法 3. 极大似然辨识算法

步骤 1. 收集高速列车的输入/合力-输出/位移的观测序列, 得到式 (4) 的辨识数据 D_N .

步骤 2. 置 $r = 0$, 初始化列车参数向量 θ_r , 设定粒子平滑的时间阈值 $Tmax$.

步骤 3. 构造列车的条件数学期望.

步骤 3.1. 执行粒子滤波算法 1, 置粒子平滑时间 $t = 0$;

步骤 3.2. 执行粒子平滑算法 2, 并记录步骤 3 的搜索时间 t , 如果 $t > Tmax$, 返回步骤 3.1, 否则得到粒子 $\{\hat{x}_k^i\}_{i=1}^M$ 及其对应的滤波权重 $\{\beta_k^i\}_{i=1}^M$ 和粒子平滑权重序列 $\{\beta_{k|N}^i\}_{i=1}^M$;

步骤 3.3. 根据式 (22) 计算 \hat{I}_1 , 根据式 (26) 计算 \hat{I}_2 , 进而由式 (27) 得到 $\hat{V}(\theta, \theta_r) = \hat{I}_1 + \hat{I}_2$.

步骤 4. 最大化列车的条件数学期望, 即根据式 (33) 更新列车参数.

步骤 5. 如果估计的列车参数满足要求 $\|\theta_{r+1} - \theta_r\| \leq 0.001$, 则算法终止, 否则转步骤 3.

4.4 辨识算法的收敛速度

EM 算法作为处理不可观测数据下的参数极大似然估计的通用方法, 对于 EM 算法的统计性质 (收敛性、一致性) 的研究也受到极大关注. 文献 [7] 和文献 [13] 探讨了 EM 算法的收敛性, 并给出了算法收敛的条件. 以此为基础, 本小节进一步分析第 4.3 节辨识算法的收敛速度.

定理 1. 如果式 (6) 定义的似然函数 $\log p_{\theta}(Y_N)$ 对于列车参数 $\theta \in \mathbf{R}^4$ 的二阶导数存在, 假设列车参数 θ^* 是似然函数 $\log p_{\theta}(Y_N)$ 局部稳定点. 那么, 根据第 4.3 节的列车辨识算法得到列车参数估计序列 θ_{r+1}, θ_r 与稳定点 θ^* 的估计误差的关系为

$$\theta_{r+1} - \theta^* = [I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)](\theta_r - \theta^*) + o(\|\theta_{r+1} - \theta_r\|) + o(\|\theta_r - \theta^*\|) \quad (34)$$

其中, $o(\cdot)$ 是高阶小量, $I \in \mathbf{R}^{4 \times 4}$ 是单位阵. 另外, $\ell_{XY}(\theta_r) \in \mathbf{R}^{4 \times 4}, \ell_Y(\theta_r) \in \mathbf{R}^{4 \times 4}$ 分别定义为

$$\ell_{XY}(\theta_r) = - \int p_{\theta_r}(x|y) \frac{\partial^2 \log p_{\theta}(x, y)}{\partial \theta \partial \theta^T} dx |_{\theta = \theta_{r+1}} = - \frac{\partial^2 V(\theta, \theta_r)}{\partial \theta \partial \theta^T} |_{\theta = \theta_{r+1}}$$

$$\ell_Y(\theta_r) = - \frac{\partial^2 \log p_{\theta}(y)}{\partial \theta \partial \theta^T} |_{\theta = \theta_r}$$

证明. 借鉴文献 [7] 和文献 [13] 的方法, 即可证明. 为节省篇幅, 在此省略. \square

说明 6. 由式 (34) 可知, 如果忽略高阶小量 $o(\|\theta_r - \theta^*\|), o(\|\theta_{r+1} - \theta_r\|)$, 则在列车参数局部稳定点 θ^* 的领域内, 由第 4.3 节辨识算法得到的参数误差可以近似为 $\theta_{r+1} - \theta^* = [I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)](\theta_r - \theta^*)$.

显然, 如果乘积项 $\ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)$ 趋于单位阵 (或 $\ell_{XY}(\theta_r) = \ell_Y(\theta_r)$), 则算法参数估计的收敛速度快. 接下来进一步分析矩阵 $\ell_{XY}^{-1}(\theta_r)$ 与 $\ell_Y(\theta_r)$ 对于列车参数估计收敛速度的影响. 为此, 根据贝叶斯定理以及列车模型 (2) 的马尔科夫性质, 可得:

$$\ell_Y(\theta_r) = \ell_{XY}(\theta_r) - \ell_X(\theta_r) \quad (35)$$

其中,

$$\ell_X(\theta_r) = - \int p_{\theta_r}(x|y) \frac{\partial \log p_{\theta}(x|y)}{\partial \theta \partial \theta^T} dx |_{\theta = \theta_r}$$

反映了不完全可测的列车状态 x 所包含的信息.

根据 $\ell_{XY}(\theta_r), \ell_Y(\theta_r), \ell_X(\theta_r)$ 的定义, 这三个矩阵均是关于列车参数向量 θ 的二阶偏导数矩阵, 从而 $\ell_{XY}(\theta_r), \ell_Y(\theta_r), \ell_X(\theta_r)$ 是对称的非负矩阵, 因此, 由式 (35) 可得:

$$\ell_Y(\theta_r) \leq \ell_{XY}(\theta_r) \quad (36)$$

由式 (36) 可知, 矩阵 $I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)$ 的所有特征值均位于 $[0, 1]$ 区间, 表明由第 4.3 节的辨识算法得到的列车参数估计误差将以指数收敛速度递减. 由式 (34) 可知误差收敛速度主要取决于矩阵 $(I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r))$ 的最大特征根 $\rho_1 = \max\{\rho[I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)]\}$, ρ_1 越小, 则参数估计误差收敛速度越快. 特别地, 如果矩阵 $\ell_{XY}(\theta_r)$ 等于 $\ell_Y(\theta_r)$, 此时 $\rho_1 = 0$, 参数估计误差将以最快的收敛速度逼近列车参数的局部稳定点.

此外, 由式 (35) 可知, 如果 $\ell_X(\theta_r)$ 取值越大 (表明不完全可测的列车状态 x 包含了更多的列车运行信息, 如列车进入隧道、列车遭遇瞬时大风等), 则矩阵 $\ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)$ 的特征值越小, 导致矩阵 $I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)$ 的特征值越大. 此时, 由第 4.3 节辨识算法得到的参数误差的收敛速度将变慢.

5 仿真说明

为了说明所提出的辨识方法用于估计高速列车参数的性能, 本文选择 CRH3 四动四拖和 CRH5 五动三拖两种动车组进行实验仿真. 其中 CRH3/CRH5 的主要特性参数^[3-4]见表 1, 牵引特性曲线^[3-4]如图 1 所示.

表 1 CRH3 和 CRH5 的主要参数特性

Table 1 The main parameters of CRH3 and CRH5

参数名称	CRH3 参数特性	CRH5 参数特性
列车总重量 (t)	500	500
最高运行速度 (km/h)	350	250
持续运营速度 (km/h)	300	200
回转质量系数	0.06	0.06
单位基本阻力 (N/kN)	$f = 0.53 + 0.0039v + 0.000114v^2$	$f = 0.693 + 0.00632v + 0.000145v^2$

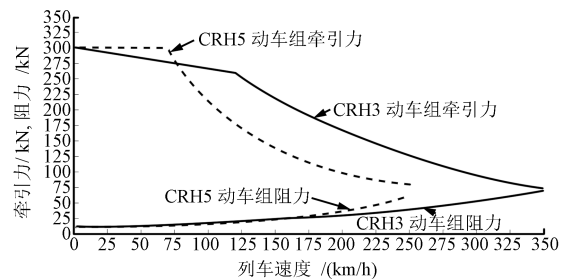


图 1 动车组 CRH3、CRH5 的牵引特性曲线
Fig. 1 Traction curves of CRH3 and CRH5

图 1 中的实线、虚线曲线分别表示 CRH3 和 CRH5 的牵引/阻力曲线. 从表 1 和图 1 可知, 两种动车组的总重均

为 500 t. 列车的回转质量系数均选取 $d = 0.06$. CRH3 的最高运行速度为 350 km/h, 持续运营速度为 300 km/h, CRH5 的最高运行速度/持续运营速度分别为 250/200 km/h. 另外, 从表 1 的最后一行可知, CRH3 和 CRH5 的单位基本阻力均是列车速度平方的非线性函数. 则 CRH3 列车需要估计的参数真值为 $\theta_{\text{CRH3}} = [0.53, 0.0039, 1.14 \times 10^{-4}, 0.06]$, CRH5 列车需要估计的参数真值为 $\theta_{\text{CRH5}} = [0.693, 0.00632, 1.45 \times 10^{-4}, 0.06]$.

基于以上列车参数特性和牵引特性曲线, 我们进行了以下两方面的数值实验.

实验 1. 进行 CRH3、CRH5 的阻力系数的估计实验, 为比较本文辨识算法对于不同车型的参数估计性能. 为此, 通过设置模型 (2) 的随机项 w_k 和 e_k 为零均值、方差为 0.01 的独立同分布的高斯白噪声来模拟列车在正常工况下的环境干扰. 以图 1 的牵引特性曲线的列车合力作为模型 (2) 的输入 u , 根据模型 (2) 分别产生 CRH3 在 300 km/h 和 CRH5 在 200 km/h 高速环境下的输出 y , 选取其中的 100 组输入-输出数据构成辨识数据集. 采用第 4.3 节的列车辨识算法进行参数估计, 辨识算法的参数设置为: 粒子数 $M = 200$, 算法迭代步数为 150, 选择列车参数真值 $(\theta_{\text{CRH3}}, \theta_{\text{CRH5}})$ 的 45% 邻域的随机数作为辨识算法的参数初始值 θ_0 . 为了降低随机噪声对列车参数估计的影响, 进行了 100 次蒙特卡洛实验 (即每次实验的随机扰动 w_k, e_k 不同, 而输入 u 固定不变).

基于上述仿真环境, 得到 CRH3 和 CRH5 实验结果分别如图 2 和图 3 所示. 图 2 和图 3 的四幅子图 (从上到下) 的纵坐标依次表示 CRH3 和 CRH5 在每次迭代中的目标函数 (式 (27) 的条件数学期望)、参数 a 、参数 b 、参数 c 所对应 100 次蒙特卡洛实验的平均值. 图 2 和图 3 的横轴标表示迭代步数.

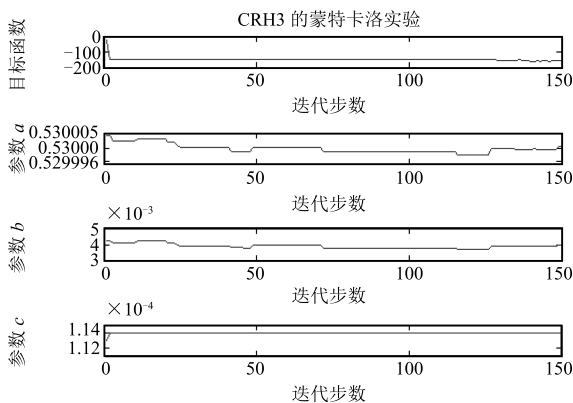


图 2 CRH3 参数辨识与迭代步数的关系

Fig. 2 The relation between average parameter estimates and iteration for CRH3

结合图 2 和图 3, 可以得出以下实验结果:

1) 从图 2 和图 3 的目标函数 (式 (27) 的条件数学期望) 可知, CRH3 的目标函数的初始值为 -23.04, 经过辨识算法的 3 步迭代, CRH3 的目标函数下降到局部最小值 -146.59, 随着迭代步数的增加, CRH3 的目标函数在第 128~132 步出现小幅波动, 在第 133 步后 CRH3 的目标函数下降到 -148.37; 类似地, CRH5 的目标函数初始值为 53.8572, 经过算法的 4 次迭代, CRH5 的目标函数下降到局部最小值 -196.28, 随着迭代步数的增加, CRH5 的目标函数一直稳定

在 -196.28. 因此, CRH3 和 CRH5 的目标函数初始值差别较大, 也即在类似的仿真环境下 CRH3 和 CRH5 对应的式 (27) 中以列车位移为条件的数学期望差异明显, 这可能是由于 CRH3 和 CRH5 不同的机械、电气等特性造成的. 此外, 辨识算法均能使 CRH3 (3 步) 和 CRH5 (4 步) 的目标函数快速地下降到局部极小值, 说明算法对于 CRH3 和 CRH5 的目标函数具有较好的优化性能.

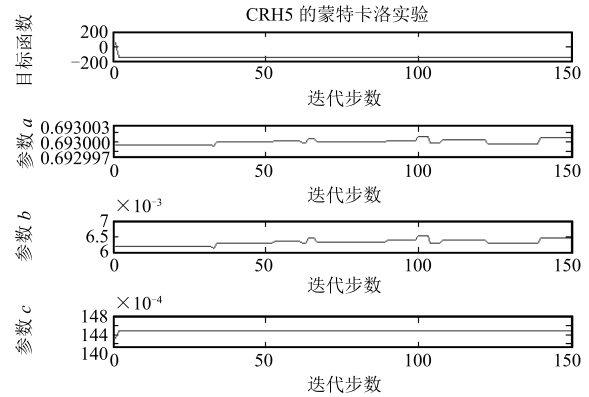


图 3 CRH5 参数辨识与迭代步数的关系

Fig. 3 The relation between average parameter estimates and iteration for CRH5

2) 从图 2 和图 3 的参数 a (列车滚动机械阻力系数) 的结果可以看出, CRH3 的参数 a 经过 25 步迭代收敛到其真值 0.53 附近. 此后, 随着辨识算法迭代的进行, CRH3 的参数 a 在第 45、50、74、109、125 步围绕其真值出现几次小幅的波动; CRH5 的参数 a 经过 2 步迭代就能逼近其真值 0.693, 随着迭代步数的增加, CRH5 的参数 a 在第 40、57、100、102、116、135 步其真值附近出现小幅波动.

3) 由图 2 和图 3 的参数 b (列车其他机械阻力系数) 的结果可知, CRH3 的参数 b 在第 26 步迭代逼近其真值 0.0039, 此后在第 49、70、125 步出现微小波动; CRH5 的参数 b 在第 38 步收敛到其真值 0.00632 附近, 此后在第 60、100、118、140 步出现小范围调整.

4) 从图 2 和图 3 的参数 c (外部空气阻力) 的结果可以得出, CRH3 经过 3 步迭代, 其参数 c 收敛到真值 0.00114 附近, 并且在此后的迭代中, CRH3 的参数 c 保持不变; CRH5 的参数 c 在第 4 步迭代逼近其真值 0.000145, 并且在随后的迭代中, CRH5 的参数 c 保持恒定.

综合以上结果可以得出, 在 CRH3 和 CRH5 的目标函数达到局部极小值后, 辨识算法迭代中获得的参数 a 和 b 的估计出现微小波动, 这是由于 EM 算法存在多个局部极小值, 这与文献 [7] 和文献 [13] 的结果是吻合的. 另外, 辨识算法对于 CRH3 和 CRH5 的参数 c 的收敛速度最快, 并且波动小, 该结果可以利用定理 1 作如下解释: 根据列车阻力方程 $f = a + bv + cv^2$, 在列车高速运行中, 与速度成平方关系的空气阻力 cv^2 占据主导地位, 由列车模型 (2) 可知 cv^2 与不完全可测的状态 x 相关. 因此, 列车速度的增加, 意味着定理 1 中的 $\ell_{XY}^{-1}(\theta_r)$ 包含了更多列车运行的信息, 进而导致式 (34) 中 $I - \ell_{XY}^{-1}(\theta_r)\ell_Y(\theta_r)$ 的最大特征值减小, 从而加快了参数 c 的收敛速度.

说明 7. 我们同时也进行了列车遭遇大风等异常工况下的数值实验. 实验中采用冲击噪声来模拟列车遭遇的大风干

扰, 即模型 (2) 的 w_k, e_k 设置为高斯分布的随机数与阶跃冲击的叠加和, 阶跃冲击的幅值根据大风的强度来确定. 具体地, 10 级大风对应的阶跃冲击幅值为 12, 8 级大风对应的阶跃冲击幅值为 9 等. 利用提出的辨识算法得到在大风环境下的 CRH3 和 CRH5 的参数估计, 对比分析大风环境与正常工况下的实验结果, 发现两者的结果类似.

实验 2. 分别采用本文方法和耦合模拟退火算法^[10] 对 CRH3 进行建模, 比较两种方法对于 CRH3 的参数优化性能. 耦合模拟退火算法 (CSA) 通过反映接受概率函数的能量和状态等信息的耦合项以实现各个模拟退火 (Simulated annealing, SA) 过程的相互耦合与信息共享, 接受函数与耦合项的逐渐匹配引导 CSA 算法获得全局最优解. 关于 CSA 的优化原理可参见文献 [10], 这里不再赘述. 采用与实验 1 类似的方法产生 CRH3 在 300 km/h 高速下位移输出, 选取其中 200 组输入-输出数据作为建模数据集. 第 4.3 节的列车辨识算法的参数设置为: 粒子数 $M = 150$, 算法迭代步数为 120, 辨识算法的初始参数选为 CRH3 参数真值的 50% 邻域的随机数. CSA 算法的参数设置为: 耦合项数为 8, 解的接受温度规则参照文献 [10] 的方法, CSA 算法的迭代次数为 120. 基于以上实验环境和算法设置, 进行了 100 次蒙特卡洛实验, 分别记录本文方法和 CSA 方法对于 CRH3 参数在 100 次实验的每步迭代的估计值, 这里给出两种方法在第 120 步迭代得到的参数 a, b, c 的 100 次实验的散点图如图 4 和图 5 所示, 具体分析如下.

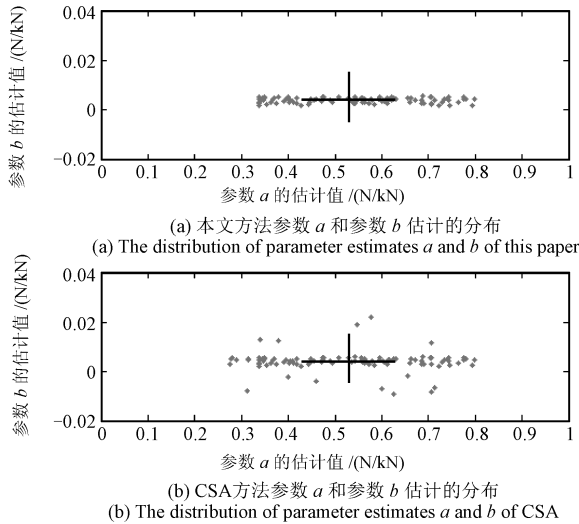


图 4 两种方法估计 a 和 b 的散点图
Fig. 4 The scatter plots of parameter estimates a and b

图 4 的横坐标和纵坐标分别表示参数 a 和参数 b 的估计值, 图 4 中两条直线的交叉点分别表示参数 (a, b) 的真值 (0.53, 0.0039), 图 4 中黑色的离散点表示 100 次蒙特卡洛实验的估计结果. 图 4(a) 和图 4(b) 分别是本文方法和 CSA 方法得到的估计结果. 从图 4(a) 可知, 在 100 次实验中, 本文方法得到的参数 a 和 b 的估计值大部分紧密分布在 a 和 b 的真值的周围, 少量实验得到的 a 和 b 的估计值出现一定幅值的偏差, 使得图 4(a) 的参数 a 的估计值的分布区间为 (0.32, 0.51), 参数 b 的估计值的分布区间为 (-0.0015, 0.0049). 从图 4(b) 可知, 在 100 次实验中, CSA 方法得到的参数 a 和 b 的估计值大部分围绕 a 和 b 的真值紧密分布, 少量实验得到的 a 和 b 的估计值出现较大偏差, 使得图 4(b) 的参数 a 的

估计值的分布范围为 (0.28, 0.81), 参数 b 的估计值的分布范围为 (-0.015, 0.22). 比较图 4(a) 和图 4(b) 可得, 本文方法得到的列车参数 a 和 b 的估计值能够更紧密地分布在 a 和 b 真值周围. CSA 方法得到的少量实验的 a 和 b 估计出现较大偏差的可能原因是: CSA 的参数优化效果受到耦合项数、解的接受温度规则、解的探究过程迭代次数等的影响^[10].

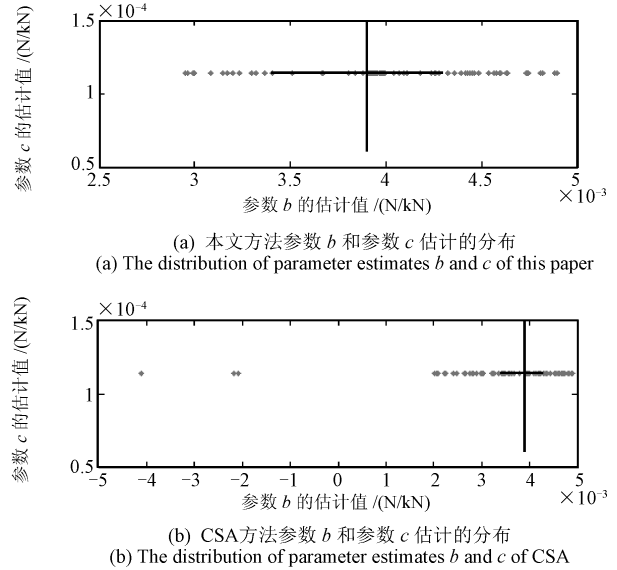


图 5 两种方法估计 b 和 c 的散点图
Fig. 5 The scatter plots of parameter estimates b and c

图 5 的横坐标和纵坐标分别表示参数 b 和参数 c 的估计值, 图 5 中两条直线的交叉点分别表示参数 (b, c) 的真值 (0.0039, 0.000114), 图 5 中黑色的离散点表示 100 次蒙特卡洛实验的估计结果. 由于在图 4 中已经对参数 b 的估计值进行了说明, 在此只对图 5 的参数 c 的估计值进行介绍. 从图 5(a) 和图 5(b) 可以看出, 两种方法得到的参数 c 的估计值均能密集分布在参数 c 的真值附近, 说明本文方法和 CSA 方法对于参数 c 的估计精度高.

为了进一步分析两种方法对于 CRH3 参数的估计性能, 表 2 给出两种方法对于 100 次蒙特卡洛实验的参数估计的均值和方差的统计结果.

表 2 两种方法得到的 100 次实验的均值与标准差
Table 2 Mean value and standard deviations of parameter estimates based on 100 Monte Carlo runs

参数	CRH3 真值	本文方法	CSA 方法
a	0.53	0.5210 ± 0.0373	0.5181 ± 0.0692
b	3.9×10^{-3}	0.0037 ± 0.00029	0.0042 ± 0.00076
c	1.14×10^{-4}	$0.0001142 \pm 2.578 \times 10^{-5}$	$0.0001141 \pm 2.63 \times 10^{-5}$
d	0.06	0.057 ± 0.0077	0.069 ± 0.0069

从表 2 可以看出, 100 次实验中, 本文方法和 CSA 方法对参数 a, b, c, d 的均值接近各自参数的真值, 说明两种方法得到的参数均值的精度高. 两种方法对于参数 a, b, c, d 估计的标准差存在差别, 本文方法对于 a 和 b 的标准差小于 CSA 方法所对应的标准差, 本文方法对于参数 d 的标准差小

于 CSA 方法的标准差, 本文方法和 CSA 方法对于参数 c 的标准差基本一致. 表 2 关于参数 a, b, c 的结果与图 4 和图 5 的对应结果吻合.

6 结论

本文提出高速列车非线性模型的极大似然 (ML) 辨识方法, 适用于具有非高斯噪声、非线性模型等复杂特征的高速列车阻力系数的在线估计. 1) 基于列车单质点力学模型的分析, 构建了高速列车的随机离散状态空间模型. 针对列车状态空间模型的状态不完全可测、非高斯噪声等特点, 将列车参数的 ML 辨识问题转化为期望极大 (EM) 的优化问题; 2) 设计了适合于列车状态条件概率密度估计的粒子滤波器和粒子平滑器, 据此构造了列车的条件数学期望, 并给出最大化列车条件数学期望的梯度搜索方法. 进而得到列车参数的辨识算法, 分析了算法的收敛速度; 3) 将提出的辨识方法应用于 CRH3 和 CRH5 的参数估计实验, 同时比较了本文方法和耦合模拟退火算法 (CSA) 对于 CRH3 的参数优化性能.

References

- Zhai Wan-Ming, Jin Xue-Song, Zhao Yong-Xiang. Some typical mechanics problems in high-speed railway engineering. *Advances in Mechanics*, 2010, **40**(4): 358–374
(翟婉明, 金学松, 赵永翔. 高速铁路工程中若干典型力学问题. 力学进展, 2010, **40**(4): 358–374)
 - Song Q, Song Y D, Tang T, Ning B. Computationally inexpensive tracking control of high-speed trains with traction/braking saturation. *IEEE Transactions on Intelligent Transportation Systems*, 2011, **12**(4): 1116–1125
 - Song Q, Song Y D. Data-based fault-tolerant control of high-speed trains with traction/braking notch nonlinearities and actuator failures. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2250–2261
 - Zhong Lu-Sheng, Yan Zheng, Yang Hui, Qi Ye-Peng, Zhang Kun-Peng, Fan Xiao-Ping. Predictive control of high-speed train based on data driven subspace approach. *Journal of the China Railway Society*, 2013, **35**(4): 77–83
(衷路生, 颜争, 杨辉, 齐叶鹏, 张坤鹏, 樊晓平. 数据驱动的高速列车子空间预测控制方法. 铁道学报, 2013, **35**(4): 77–83)
 - Zhong Lu-Sheng, Yan Zheng, Gong Jin-Hong, Zhang Yong-Xian, Zhu Zhen-Min, Fan Xiao-Ping. Adaptive subspace predictive control of high-speed train based on time-varying forgetting factor. *Journal of the China Railway Society*, 2013, **35**(5): 54–61
(衷路生, 颜争, 龚锦红, 张永贤, 祝振敏, 樊晓平. 时变遗忘因子的高速列车自适应子空间预测控制. 铁道学报, 2013, **35**(5): 54–61)
 - Zhong Lu-Sheng, Yan Zheng, Gong Jin-Hong, Zhang Yong-Xian, Zhu Zhen-Min. Adaptive subspace predictive control method of high-speed train. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2013, **41**(8): 28–35
(衷路生, 颜争, 龚锦红, 张永贤, 祝振敏. 高速列车的自适应子空间预测控制方法. 华中科技大学学报, 2013, **41**(8): 28–35)
 - Gibson S, Ninness B. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 2005, **41**(10): 1667–1682
 - Zhong Lu-Sheng, Song Zhi-Huan. Hierarchical optimization identification of LTI state-space by projected gradient search *Acta Automatica Sinica*, 2008, **34**(6): 711–715
(衷路生, 宋执环. 基于正交梯度搜索的动态系统递阶优化辨识. 自动化学报, 2008, **34**(6): 711–715)
 - Zhong Lu-Sheng, Fan Xiao-Ping, Yang Hui, Qu Zhi-Hua. Output error identification of LTI state-space models by orthogonal gradient search. *Control and Decision*, 2011, **26**(5): 685–689
(衷路生, 樊晓平, 杨辉, 瞿志华. 状态空间模型基于正交梯度搜索的预报误差辨识. 控制与决策, 2011, **26**(5): 685–689)
 - De Brabanter K. Least Squares Support Vector Regression with Applications to Large-Scale Data: a Statistical Approach [Ph. D. dissertation], Katholieke Universiteit Leuven, Faculty of Engineering, Belgium, 2011
 - Schon T B, Wills A, Ninness B. System identification of nonlinear state-space models. *Automatica*, 2011, **47**: 39–49
 - Cheng Yu-Hu, Feng Huan-Ting, Wang Xue-Song. Expectation-maximization policy search with parameter-based exploration. *Acta Automatica Sinica*, 2012, **38**(1): 38–45
(程玉虎, 冯涣婷, 王雪松. 基于参数探索的期望最大化策略搜索. 自动化学报, 2012, **38**(1): 38–45)
 - Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977, **39**(1): 1–38
 - Zhang Gong-Yuan, Cheng Yong-Mei, Yang Feng, Pan Quan, Liang Yan. Design of an adaptive particle filter based on variance reduction technique. *Acta Automatica Sinica*, 2010, **36**(7): 1020–1024
- 衷路生** 华东交通大学电气与电子工程学院副教授. 主要研究方向为系统辨识, 统计学习理论. 本文通信作者.
E-mail: lszhongzju@gmail.com
(ZHONG Lu-Sheng Associate professor at the School of Electrical and Electronic Engineering, East China Jiaotong University. His research interest covers system identification and statistical learning. Corresponding author of this paper.)
- 李兵** 华东交通大学电气与电子工程学院硕士研究生. 主要研究方向为系统辨识, 统计学习理论. E-mail: libing87@gmail.com
(LI Bing Master student at the School of Electrical and Electronic Engineering, East China Jiaotong University. His research interest covers system identification and statistical learning.)
- 龚锦红** 华东交通大学电气与电子工程学院讲师. 主要研究方向为控制理论, 统计学习理论. E-mail: jhgong@ecjtu.jx.cn
(GONG Jin-Hong Lecturer at the School of Electrical and Electronic Engineering, East China Jiaotong University. His research interest covers control theory and statistical learning.)
- 张永贤** 华东交通大学电气与电子工程学院副教授. 主要研究方向为网络控制, 优化理论. E-mail: yxzhang@ecjtu.jx.cn
(ZHANG Yong-Xian Associate professor at the School of Electrical and Electronic Engineering, East China Jiaotong University. His research interest covers network control and optimization theory.)
- 祝振敏** 华东交通大学电气与电子工程学院讲师. 主要研究方向为光学理论, 系统建模. E-mail: zmzhu@ecjtu.jx.cn
(ZHU Zhen-Min Lecturer at the School of Electrical and Electronic Engineering, East China Jiaotong University. His research interest covers optic theory and system modeling.)