Asia Pac. J. Math. 2026 **13**:183

CONSTRAINED OPTIMAL CONTROL OF VARIABLE-ORDER CAPUTO SYSTEMS: CONDITIONAL CONVERGENCE RATES FOR A SHIFTED LEGENDRE TRANSCRIPTION AND THE REGULARITY BARRIER

KHADIGA WADI NAHAR TAJER[1], MANAL YAGOUB JUMA[1], MANAHIL A. M. ASHMAIG[1], NIDAL E. TAHA[1,*], BATUL ALI ALBALULAH MAHMOUD[2], HALA MOHAMMED ELHASSAN AHMED[2], THWIBA A. KHALID[3,4]

[1]Department of Mathematics, College of Science, Qassim University, Buraidah 51452, Saudi Arabia
[2]Department of Mathematics, College of Sciences, King Khalid University, Abha 61421, Saudi Arabia
[3]Department of Mathematics, Faculty of Science, Al-Baha University, Al-Baha 65779, Saudi Arabia
[4]Department of Mathematics, College of Engineering and Medical Sciences, Khartoum 11111, Sudan
[*]Corresponding author: n.taha@qu.edu.sa

Received Sep. 9, 2026

Abstract. Shifted Legendre transcription is by now an established route to variable-order (VO) fractional optimal control problems: the state and control are expanded in a global polynomial basis, the VO Caputo derivative is realised through a node-dependent differentiation matrix, and the problem is reduced to a nonlinear programme. Two things, however, are missing from the existing treatments. First, the problems solved are unconstrained — no actuator bounds, no pointwise state restrictions — although these are precisely what make a control problem practical. Second, the convergence results available assume that the discrete solutions converge and then verify that the limit is optimal; no rate is established, and the empirical claim of spectral accuracy is documented almost exclusively at integer order. This paper addresses both. We formulate the constrained VO-FOCP, impose control and state constraints, and bound how far a discretely feasible solution can violate them in between, so that a matching constraint tightening restores feasibility on the whole horizon. We then prove an error estimate $\|x - x_N\|_{L^2} = O(N^{2-m})$ for states in H^m, together with $|J_N^* - J^*| = O(N^{2-m})$ and subsequential convergence of the discrete minimisers, using a variable-order Grönwall inequality that we state and prove. These rates are *conditional*: they rest on an operator-stability hypothesis (A5), a stability estimate for the variable-exponent Abel operator, which we prove only under a smallness condition on the oscillation of α, the general case being open. Finally we identify an obstruction that these rates conceal. The solution of a VO initial value problem carries a $t^{\alpha(0)}$ boundary layer, and we show — analytically and by direct computation of the best L^2 polynomial approximation — that the attainable rate saturates at $N^{-(2\alpha(0)+1)}$, so spectral accuracy in the usual sense is not available at genuinely fractional order. A p-term singular enrichment is proved to raise the algebraic exponent, and the gain is confirmed numerically. Four numerical experiments are reported, including

DOI: 10.28924/APJM/13-183

constrained optimal regulation of a variable-order Arneodo system, contrasted with the order-modulation controller of our earlier work.

2020 Mathematics Subject Classification. 26A33; 34A08; 34H10; 41A25; 45E10; 46E35; 49J15; 49M25; 65L60; 93C10.

Key words and phrases. variable-order fractional calculus; Caputo derivative; variable-exponent Abel operator; weakly singular Volterra equation; constrained optimal control; shifted Legendre collocation; conditional convergence rate; Sobolev regularity; algebraic saturation; regularity barrier; chaos control.

1. Introduction

Fractional operators encode memory that classical derivatives cannot represent. Allowing the order itself to vary adds a further degree of realism: introduced by Samko and Ross [1] and developed by Lorenzo and Hartley [2] and Coimbra [3], variable-order (VO) operators describe media whose memory structure is not stationary — a transport regime drifting from subdiffusive to normal, a material whose relaxation spectrum follows its loading history, a dynamical system whose dissipation is modulated externally. The surveys [4,5] document the range of these applications.

The price is analytical. The VO Caputo derivative ${}^{C}D_{0^+}^{\alpha(t)}$ has no semigroup property, no closed-form Laplace transform for general $\alpha(\cdot)$, and — decisively for control — no fractional integration-by-parts identity of the kind that underlies Pontryagin conditions at constant order. The adjoint of ${}^{C}D_{0^+}^{\alpha(t)}$ is not a right-sided VO derivative of the same profile, and the manipulations that close for constant α [6–8] do not close here. Even well-posedness of the state equation requires separate treatment [9–11].

1.1. What the literature has established.

The response has been overwhelmingly numerical. Wavelets [12], generalized polynomials [13,14], generalized Bessel polynomials [15], piecewise schemes [16] and fast iterative solvers for the coupled forward–backward system [17,18] have all been applied to VO fractional optimal control problems (VO-FOCPs). Most recently, Pirouzeh, Nassiri Pirbazari and Noori Skandari [19] proposed a shifted Legendre spectral collocation (SLSC) scheme built on the operational matrix $\mathbf{D}^{\{\rho(t)\}} = \mathbf{Q}\,\mathbf{Y}^{\{\rho(t)\}}\mathbf{Q}^{-1}$ evaluated at each quadrature node, reduced the VO-FOCP to a nonlinear programme, and reported six numerical examples.

We take [19] as our starting point rather than as a competitor, and we state plainly which parts of the present construction are shared with it: the basis, the node-dependent realisation of the VO derivative, and the reduction to an NLP are, in substance, the same. Our contribution lies in two directions in which [19] and the wider VO-FOCP literature are silent.

(i) Constraints. The problem treated in [19] carries no control bounds and no state restrictions; its concluding section lists state constraints explicitly among the open problems. This is not a minor omission. Actuator saturation is the norm rather than the exception in applications, and once constraints are imposed at a finite set of collocation nodes, feasibility between nodes is no longer automatic — a question that does not arise at all in the unconstrained setting.

(ii) A rate, under a stated hypothesis. The convergence theorem of [19] (their Theorem 2) assumes that the sequence of discrete solutions converges uniformly to a limit, and then shows the limit is optimal. That is a consistency statement, and a useful one, but it is not a convergence proof: it establishes nothing about whether the sequence converges, nor how fast. No error estimate in N appears. The accompanying claim of exponential convergence is supported by error tables computed at $\rho(t) = 1$ — the integer-order case — with analytic exact solutions (t^2, e^{-t}, t^3). Whether spectral accuracy survives at genuinely fractional order is left untested. At integer order both kinds of statement are available for pseudospectral transcription: the method originates with [20,21], and convergence rates for Gauss and Radau collocation, unconstrained and constrained, are established in [22–24]. No counterpart exists at variable order, and supplying one is the second aim of the present paper.

Solutions of VO initial value problems generically behave like $x(t) = x_0 + c\,t^{\alpha(0)} + o(t^{\alpha(0)})$ near the origin, so $x \in H^m$ only for $m < \alpha(0) + 1/2$. We show that the best L^2 polynomial approximation of $t^{\alpha(0)}$ decays exactly like $N^{-(2\alpha(0)+1)}$ — algebraically, not exponentially — and we verify this numerically to three decimal places (Section 7, Figure 6). An exponential rate measured at $\alpha \equiv 1$ is therefore a statement about the integer-order case; it does not establish a rate for the variable-order problem, where the boundary layer is present by construction.

1.2. **Relation to our previous work.** In [25] we developed shifted Legendre collocation for VO fractional advection–dispersion equations: a forward problem, with no control variable, cost functional or constraints. The algebraic apparatus there is the starting point for Section 4.

In [26] we studied a VO Arneodo system and used order modulation — a prescribed profile $\alpha(t)$ — to suppress chaos and achieve hybrid synchronisation, with stability assessed through Matignon's criterion. That study contains no optimality criterion: the profile is designed heuristically, actuation effort is unpenalised, no bound restricts the admissible control, and no convergence guarantee accompanies the scheme. Actuator saturation and bounded disturbances for constant-order fractional chaotic plants have been addressed in [27–29], but in a Lyapunov-design rather than an optimal-control setting.

The present paper departs from both: an explicit quadratic cost trades tracking against control energy, saturation and state restrictions are part of the problem statement, and the discretisation comes with a proved rate.

1.3. **What is new, and what is not.** Because this paper builds on [19] and on two of our own previous studies, we state the boundary explicitly rather than leaving it to be inferred.

Taken over unchanged from [19]: the shifted Legendre basis; the idea of realising the VO Caputo derivative through a node-dependent matrix; the reduction of the control problem to a nonlinear programme. We claim no novelty for any of these.

Taken over from our own work: the algebraic construction of the differentiation matrix, developed in [25] for a forward problem with no control variable, cost functional or constraints; and the VO Arneodo plant and its order profile, studied in [26] with a heuristically designed order modulation, no optimality criterion, no effort penalty and no input bound.

New here:

(N1) The constrained formulation itself. Control bounds and pointwise state constraints do not appear in [19], which lists state constraints among its open problems, nor in [25], which has no control variable.

(N2) Proposition 5.2: imposing constraints at the collocation nodes is unsafe, and the worst-case between-node violation grows like $\log N$. We have not found this observed in the fractional optimal control literature.

(N3) Proposition 5.3 and the accompanying tightening: a constraint grid decoupled from the collocation grid, with an N-independent overshoot bound.

(N4) Lemma 6.1, a Grönwall inequality adapted to the non-convolution VO kernel, and Theorems 6.3–6.4, which give a rate where [19] gives consistency under an assumption of convergence. The rates are conditional on (A5); see below.

(N5) Lemma 3.2, which isolates (A5) and proves it for small oscillation of α. This replaces a stronger unproved hypothesis by a weaker one that is proved in a nontrivial regime.

(N6) Theorems 7.1 and 7.2 and Proposition 7.4: the regularity barrier, the enrichment that raises the exponent, and the residual floor at the origin. The saturation exponent (22) is classical for Jacobi projection; what is new is drawing its consequence for VO control and quantifying it in the control setting.

(N7) The conditioning analysis of Remark 4.3, measured against an extended-precision reference, and the resulting practical ceiling $N \approx 24$.

(N8) Remark 5.1: enrichment leaves the transcription under-determined unless the node count is raised.

1.4. **Contributions.**

(1) A constrained VO-FOCP formulation with control bounds and pointwise state constraints, together with an existence result (Proposition 3.5) and a self-contained well-posedness proof for the state equation (Proposition 3.4) that avoids the integral-equation reformulation unavailable at variable order.

(2) A between-node feasibility analysis with a negative and a positive half (Section 5.1). Constraints imposed at the $N+1$ collocation nodes can be violated between them by the Lebesgue constant of the node set, $\sim \frac{2}{\pi} \ln N$, so node-wise imposition is not safe and degrades under refinement.

Imposing them instead on a Chebyshev grid of $\kappa(N+1)$ points bounds the overshoot by $1+O(\kappa^{-2})$, independently of N.

(3) A variable-order Grönwall inequality (Lemma 6.1, proved in Appendix A), and, on its basis, a quantitative error estimate $\|x-x_N\|_{L^2}=O(N^{2-m})$ (Theorem 6.3), convergence of the optimal cost $|J_N^*-J^*|=O(N^{2-m})$ and subsequential convergence of minimisers (Theorem 6.4).

(4) Identification and quantification of the regularity barrier (Theorem 7.1): the attainable rate saturates at $N^{-(2\alpha(0)+1)}$, and for a solution of the form assumed in Theorem 7.2 a p-term singular enrichment raises the exponent from $2a+1$ to $2b+1$; with $b=a+\frac{1}{2}$ this is a gain of one, and the measured gains are 1.22 and 1.24 (Figure 6). We also record a consequence of the barrier that is visible in a quantity practitioners actually check: since the Caputo image of any polynomial vanishes at the origin while a solution satisfies ${}^{C}D_{0+}^{\alpha}x(t)\to f(0,x_0,u(0))$, the continuous residual near $t=0$ is bounded below by $|f(0,x_0,u(0))|$ for every N (Proposition 7.4).

(5) Four numerical experiments (Section 8), including constrained optimal regulation of the VO Arneodo system compared directly with the order-modulation controller of [26], and a publicly archived Python implementation with fully specified benchmarks.

2. Preliminaries

Throughout $T>0$ is fixed, $I=[0,T]$, and $\alpha: I\to(0,1]$ is prescribed.

2.1. The variable-order Caputo derivative.

Definition 2.1. For $x\in AC(I)$ and $0<\alpha(t)\le 1$,

$$ {}^{C}D_{0+}^{\alpha(t)}x(t)=\frac{1}{\Gamma(1-\alpha(t))}\int_0^t (t-s)^{-\alpha(t)}\,x'(s)\,ds,\qquad t\in(0,T], \tag{1}$$

with ${}^{C}D_{0+}^{1}x=x'$ wherever $\alpha(t)=1$.

2.2. Which variable-order operator?

Equation (1) freezes the exponent at the evaluation point. At least three inequivalent conventions circulate, and results obtained under different conventions are not comparable. Table 1 records the distinction, which is rarely made explicit. That the choice is not cosmetic has been shown directly: in [30] the validity of the quadratic Lyapunov inequality for a variable-order derivative is found to turn on whether the memory kernel is monotone in the integration variable, a property that holds for some of these conventions and fails for others, so that a stability conclusion drawn under one definition need not survive a change of definition.

The first convention is the only one for which the action on monomials is elementary, which is why every operational-matrix construction — including [12, 19] and the present one — adopts it. The third is the only convolution operator of the three and is therefore the only one admitting fast summation;

TABLE 1. Three variable-order Caputo conventions. Only the first is used here.

Kernel exponent	Expression	Kernel	Behaviour on t^k
frozen at t (used here)	$\dfrac{1}{\Gamma(1-\alpha(t))}\displaystyle\int_0^t (t-s)^{-\alpha(t)}x'(s)ds$	non-Toeplitz	closed form (2)
evaluated at s	$\displaystyle\int_0^t \dfrac{(t-s)^{-\alpha(s)}}{\Gamma(1-\alpha(s))}x'(s)ds$	non-Toeplitz	no closed form
evaluated at $t-s$	$\displaystyle\int_0^t \dfrac{(t-s)^{-\alpha(t-s)}}{\Gamma(1-\alpha(t-s))}x'(s)ds$	convolution	no closed form

the price is that no closed-form image of t^k exists. This trade-off deserves to be stated rather than passed over.

Lemma 2.2. *For $k \in \mathbb{N}$ and $0 < \alpha(t) < 1$,*

$$ {}^{C}D_{0^+}^{\alpha(t)} t^k = \begin{cases} 0, & k = 0, \\ \dfrac{\Gamma(k+1)}{\Gamma(k+1-\alpha(t))}\, t^{\,k-\alpha(t)}, & k \geq 1. \end{cases} \tag{2} $$

Proof. For $k = 0$ the integrand vanishes. For $k \geq 1$, substituting $x'(s) = ks^{k-1}$ in (1) and setting $s = t\sigma$,

$$ {}^{C}D_{0^+}^{\alpha(t)} t^k = \frac{k\, t^{k-\alpha(t)}}{\Gamma(1-\alpha(t))} \int_0^1 (1-\sigma)^{-\alpha(t)} \sigma^{k-1} d\sigma = \frac{k\, t^{k-\alpha(t)} B(k, 1-\alpha(t))}{\Gamma(1-\alpha(t))}, $$

and the Beta–Gamma identity gives (2). □

Remark 2.3 (No semigroup property). For non-constant α one has, in general, ${}^{C}D_{0^+}^{\alpha(t)}\,{}^{C}D_{0^+}^{\beta(t)} x \neq {}^{C}D_{0^+}^{\alpha(t)+\beta(t)} x$, and the composition of ${}^{C}D_{0^+}^{\alpha(t)}$ with the corresponding VO Riemann–Liouville integral is not the identity. The equivalence between a fractional differential equation and a Volterra integral equation, routine at constant order, is therefore unavailable, and the standard existence proof by contraction on the integral form does not apply. Proposition 3.4 works directly with the differential form instead. The same obstruction removes the transform-based route: decomposition methods built on an integral transform, effective at constant order — see for instance the Elzaki transform decomposition of [31] for time-fractional evolution equations — have no analogue here, because the kernel of (1) is not of convolution type and admits no closed-form transform for general $\alpha(\cdot)$.

Figure 1 illustrates what is at stake. The kernel of (1) is not a convolution: its exponent changes with the evaluation time, so the operator has a different memory profile at every instant. The right panel shows the discrepancy incurred by replacing $\alpha(\cdot)$ with a representative constant — a substitution that converts the problem into a different one.

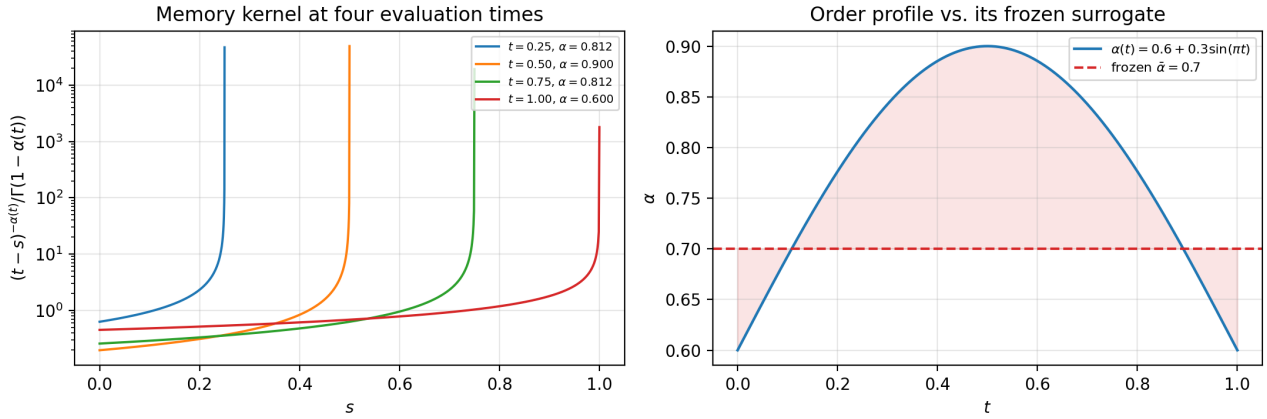


FIGURE 1. Left: the memory kernel $(t-s)^{-\alpha(t)}/\Gamma(1-\alpha(t))$ of the VO Caputo operator at four evaluation times, for $\alpha(t) = 0.6 + 0.3\sin(\pi t)$. Both the strength and the exponent of the singularity change with t, so the operator is not a convolution and no Toeplitz structure is available. Right: the order profile against the frozen surrogate $\bar{\alpha} = 0.7$; the shaded area is the modelling error committed by freezing.

2.3. **Shifted Legendre polynomials.** Let P_i be the Legendre polynomial of degree i on $[-1, 1]$ and $\varphi_i(t) = P_i(2t/T - 1)$ the shifted family on I, so that

$$\int_0^T \varphi_i \varphi_j \, dt = \frac{T}{2i+1}\delta_{ij}, \qquad \varphi_i(t) = \sum_{k=0}^{i} c_{ik} t^k, \qquad c_{ik} = \frac{(-1)^{i+k}(i+k)!}{(i-k)!(k!)^2 T^k}. \tag{3}$$

Write $\mathcal{P}_N = \operatorname{span}\{\varphi_0, \ldots, \varphi_N\}$ and let π_N be the L^2 orthogonal projection onto it. If $x \in H^m(I)$, $m \geq 1$, then

$$\|x - \pi_N x\|_{L^2} \leq C_1 N^{-m} \|x^{(m)}\|_{L^2}, \qquad \|x - \pi_N x\|_{L^\infty} \leq C_2 N^{1/2-m} \|x^{(m)}\|_{L^2}. \tag{4}$$

Collocation uses the $N+1$ shifted Legendre–Gauss–Radau (LGR) nodes $0 = t_0 < t_1 < \cdots < t_N < T$. We depart here from [19], which uses Gauss–Lobatto nodes. LGR includes the left endpoint — where the initial condition lives and where the kernel is singular — but excludes the right endpoint, which avoids collocating the dynamics at a point where the polynomial approximation is least constrained. The distinction is minor for smooth data and material when a boundary layer is present.

3. The Constrained Problem

We consider

$$\min_{u \in \mathcal{U}_{\mathrm{ad}}} \; J(x,u) = \int_0^T \Big[x^\top Q x + u^\top R u \Big] dt + \Phi(x(T)) \tag{5}$$

subject to

$${}^{C}D_{0^+}^{\alpha(t)} x(t) = f(t, x(t), u(t)), \quad t \in (0,T], \qquad x(0) = x_0, \tag{6}$$

$$u_{\min} \leq u(t) \leq u_{\max}, \qquad g(x(t)) \leq 0, \qquad \text{a.e. } t \in I, \tag{7}$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^r$, $Q \succeq 0$, $R \succ 0$, $\Phi \in C^1(\mathbb{R}^n; \mathbb{R}_{\geq 0})$, $g \in C^1(\mathbb{R}^n; \mathbb{R}^q)$, and

$$\mathcal{U}_{\mathrm{ad}} = \{u \in L^\infty(I; \mathbb{R}^r) : u_{\min} \leq u(t) \leq u_{\max} \text{ a.e.}\}.$$

Setting $u_{\min} = -\infty$, $u_{\max} = +\infty$ and $g \equiv -1$ recovers the unconstrained formulation of [19], so everything below applies to that case as a special instance.

Assumption 3.1.

(A1) f is continuous and Lipschitz in (x, u) with constant L_f on a bounded set containing the reachable states.

(A2) $\alpha \in C^{0,\gamma}(I)$ for some $\gamma \in (0, 1]$ and $0 < \alpha_- \leq \alpha(t) \leq \alpha_+ < 1$ on I.

(A3) The feasible set of (5)–(7) is nonempty.

(A4) g is Lipschitz with constant L_g on the reachable set.

(A5) (*Operator stability.*) There is $C_S > 0$ such that every $\theta \in AC(I)$ with $\theta(0) = 0$ satisfies $\|\theta\|_{L^2(I)} \leq C_S \big\| {}^C D_{0^+}^{\alpha(\cdot)} \theta \big\|_{L^2(I)}$.

Hypotheses (A1)–(A4) are standard. (A5) is not, and it deserves comment rather than silence, because it is the hinge on which every convergence claim for a variable-order collocation scheme turns.

At constant order α, (A5) is a theorem, not an assumption: since $\theta(0) = 0$ one has $\theta = I^\alpha \, {}^C D_{0^+}^{\alpha} \theta$ exactly, whence $C_S = T^\alpha / \Gamma(\alpha + 1)$. At variable order this inversion is not available. Remark 2.3 is precisely the statement that $I^{\alpha(\cdot)} {}^C D_{0^+}^{\alpha(\cdot)} \neq \mathrm{Id}$: the operator (1) is a Volterra operator with kernel $k(t, s) = (t - s)^{-\alpha(t)} / \Gamma(1 - \alpha(t))$ which, the exponent depending on the outer variable, is not of convolution type, and its inverse is not a fractional integral of any order. Abel's integral equation and its inversion go back to Abel's own treatment of the tautochrone, and are developed systematically in [32]. What (A5) asks for is not that inversion — the Abel integral operator is compact, and inverting it is the classical ill-posed problem — but a lower bound for ${}^C D_{0^+}^{\alpha(\cdot)}$ on $\{\theta(0) = 0\}$, equivalently boundedness of the solution operator of ${}^C D_{0^+}^{\alpha(\cdot)} \theta = f$. At constant order the two coincide, because that solution operator is I^α; at variable order it has no closed form. Lemma 3.2 below establishes (A5) under a smallness condition on the oscillation of α; the general case is open, and we say so rather than assuming it away.

Lemma 3.2 (Sufficient condition for (A5))**.** *Let (A2) hold and set* $\alpha^* = \frac{1}{2}(\alpha_- + \alpha_+)$ *and* $\omega = \alpha_+ - \alpha_-$. *Define the perturbation operator*

$$(E\theta)(t) = \int_0^t \left[\frac{(t - s)^{-\alpha(t)}}{\Gamma(1 - \alpha(t))} - \frac{(t - s)^{-\alpha^*}}{\Gamma(1 - \alpha^*)} \right] \theta'(s) \, ds.$$

If E *is bounded from* $\{\theta \in H^1 : \theta(0) = 0\}$ *to* $L^2(I)$ *with* $\|I^{\alpha^*} E\| < 1$ *as an operator on* $\{\theta \in L^2 : \theta(0) = 0\}$ — *which holds whenever*

$$\omega \left(1 + |\ln T| + \frac{1}{1 - \alpha_+} \right) \frac{T^{\alpha^*}}{\Gamma(\alpha^* + 1)} < 1, \tag{8}$$

then (A5) holds with $C_S = T^{\alpha^*} \Gamma(\alpha^* + 1)^{-1} \big(1 - \|I^{\alpha^*} E\|\big)^{-1}$.

Proof. Write ${}^{C}D_{0+}^{\alpha(t)}\theta = {}^{C}D_{0+}^{\alpha^*}\theta + E\theta$. Since $\theta(0) = 0$, the constant-order inversion is legitimate: $\theta = I^{\alpha^*}{}^{C}D_{0+}^{\alpha^*}\theta$. Substituting gives $(\mathrm{Id} + I^{\alpha^*}E)\theta = I^{\alpha^*}{}^{C}D_{0+}^{\alpha(\cdot)}\theta$, and the Neumann series converges under the stated norm bound. The sufficient condition (8) follows from $|(t-s)^{-a} - (t-s)^{-b}| \le |a-b|(t-s)^{-\max(a,b)}|\ln(t-s)|$ together with the bound $\|I^{\alpha^*}\|_{L^2\to L^2} \le T^{\alpha^*}/\Gamma(\alpha^*+1)$ and integrability of $(t-s)^{-\alpha_+}|\ln(t-s)|$. □

Remark 3.3 (What is and is not proved). Condition (8) is restrictive: for $T = 1$ and $\alpha_+ = 0.9$ it requires $\omega \lesssim 0.08$, whereas the profile $\alpha(t) = 0.6 + 0.3\sin(\pi t)$ used in our experiments has $\omega = 0.3$. We therefore do not claim that Theorems 6.3 and 6.4 are unconditional; they hold under (A5), which we can verify only for mildly varying order.

We nevertheless regard this as progress rather than a retreat. Every convergence statement in the variable-order collocation literature requires something of this kind, and the alternatives are weaker: the result of [19] assumes uniform convergence of the discrete solutions outright, which is strictly stronger than (A5) and is the conclusion one wants to prove. Isolating the exact operator-theoretic property on which the analysis depends, proving it in a nontrivial regime, and stating plainly that the general case is open seems to us more useful than a theorem whose hypotheses conceal the difficulty. Establishing (A5) for arbitrary Hölder α — equivalently, a stability estimate for the variable-exponent Abel operator — is, in our view, the outstanding analytical question in this subject, alongside the variable-order Pontryagin principle.

Proposition 3.4 (Well-posedness of the state equation)**.** *Under (A1)–(A2), for every $u \in \mathcal{U}_{\mathrm{ad}}$ problem (6) has a unique solution $x[u] \in C(I;\mathbb{R}^n) \cap AC(I;\mathbb{R}^n)$, and the solution map is Lipschitz from $L^2(I;\mathbb{R}^r)$ to $C(I;\mathbb{R}^n)$.*

Sketch. Remark 2.3 blocks the usual route through the integral equation, so we argue on the derivative. Put $v = x'$ and rewrite (6) as the weakly singular Volterra equation of the second kind

$$\int_0^t (t-s)^{-\alpha(t)} v(s)\,ds = \Gamma(1-\alpha(t))\, f\!\left(t,\, x_0 + \int_0^t v,\, u(t)\right), \tag{9}$$

which is a genuine equation for v because the left-hand side is a bounded injective operator on $L^1(I)$ for each fixed t under (A2). Introduce on $C(I)$ the weighted norm $\|y\|_\lambda = \sup_{t\in I} e^{-\lambda t}|y(t)|$ and let $\mathcal{T}$ denote the map sending y to the right-hand side of (9) with $x = x_0 + \int_0^{\cdot} y$. Using (A1) and the uniform bound $\int_0^t (t-s)^{-\alpha(t)} e^{\lambda s}ds \le e^{\lambda t}\lambda^{\alpha_+-1}\Gamma(1-\alpha_+)$, one obtains $\|\mathcal{T}y_1 - \mathcal{T}y_2\|_\lambda \le L_f\Gamma(1-\alpha_-)\lambda^{\alpha_+-1}\Gamma(1-\alpha_+)\|y_1 - y_2\|_\lambda$, and since $\alpha_+ < 1$ the factor $\lambda^{\alpha_+-1} \to 0$ as $\lambda \to \infty$. Choosing λ large makes $\mathcal{T}$ a contraction; Banach's theorem gives existence and uniqueness, and the same estimate applied to two controls yields the Lipschitz dependence. □

Proposition 3.5 (Existence of a minimiser)**.** *Under Assumption 3.1 and Proposition 3.4, and if the feasible set is weakly closed in $L^2(I;\mathbb{R}^r)$, problem (5)–(7) admits at least one solution (x^*, u^*).*

Proof. $\mathcal{U}_{\mathrm{ad}}$ is bounded, convex and closed in L^2, hence weakly sequentially compact. Take a minimising sequence $u_k \rightharpoonup \bar{u} \in \mathcal{U}_{\mathrm{ad}}$. By Proposition 3.4, $x[u_k] \to x[\bar{u}]$ uniformly along a subsequence; feasibility of $\bar{u}$ follows from (A4) and pointwise passage to the limit. The integrand is convex in u and continuous in x, so J is weakly lower semicontinuous and $J(x[\bar{u}], \bar{u}) \le \liminf_k J(x[u_k], u_k) = J^*$. □

Uniqueness is not claimed: f is not assumed affine, and Example 8.4 involves a chaotic plant for which the problem is manifestly nonconvex. We also do not invoke Mittag-Leffler stability [33] as a design tool, since its extension to non-constant order is not established: [30] exhibits a variable-order operator for which even the quadratic Lyapunov inequality fails, together with sufficient conditions, in terms of kernel monotonicity, under which it is recovered. Closed-loop behaviour is therefore assessed a posteriori from computed trajectories.

4. Node-Dependent Differentiation Matrix

Write $\Phi_N(t) = [\varphi_0(t), \dots, \varphi_N(t)]^\top$. From (3) and Lemma 2.2,

$$ {}^{C}D_{0^+}^{\alpha(t)} \varphi_i(t) = \sum_{k=1}^{i} c_{ik} \frac{\Gamma(k+1)}{\Gamma(k+1-\alpha(t))} t^{k-\alpha(t)}. \tag{10} $$

The nodal matrix used throughout is

$$ [\mathbf{D}_N]_{ji} = {}^{C}D_{0^+}^{\alpha(t)} \varphi_i(t)\Big|_{t=t_j} = \sum_{k=1}^{i} c_{ik} \frac{\Gamma(k+1)}{\Gamma(k+1-\alpha(t_j))} t_j^{k-\alpha(t_j)}, \tag{11} $$

together with $[\mathbf{V}_N]_{ji} = \varphi_i(t_j)$. Equation (11) is the entrywise form of the operational matrix $\mathbf{Q}\mathbf{Y}^{\{\alpha(t_j)\}}\mathbf{Q}^{-1}$ of [19] and is computed here directly, without forming $\mathbf{Q}^{-1}$, for the conditioning reasons given in Remark 4.3.

Proposition 4.1. *A matrix $D^{(\alpha)}$ satisfying ${}^{C}D_{0^+}^{\alpha(t)} \Phi_N(t) = D^{(\alpha)} \Phi_N(t)$ for all $t \in I$ exists if and only if α is constant on I.*

Proof. If $\alpha \equiv \alpha_0$, (10) is a fixed combination of $t^{k-\alpha_0}$ and the classical construction applies. Conversely, taking $i = 1$ gives ${}^{C}D_{0^+}^{\alpha(t)} \varphi_1(t) = (2/T) t^{1-\alpha(t)} / \Gamma(2-\alpha(t))$, which would have to equal a fixed polynomial of degree $\le N$. A function $t^{1-\alpha(t)}\psi(t)$ with ψ continuous and positive cannot agree with a polynomial on a nondegenerate interval unless $1 - \alpha(t)$ is constant there. □

Proposition 4.1 is elementary, and implementations that evaluate the matrix node by node — as [19] does — already respect it in practice. We record it because the phrase "operational matrix for the variable-order derivative" is used in the literature for both the node-dependent object and for constant matrices obtained by freezing α at a representative value, and the two are not interchangeable: the second silently solves a constant-order problem.

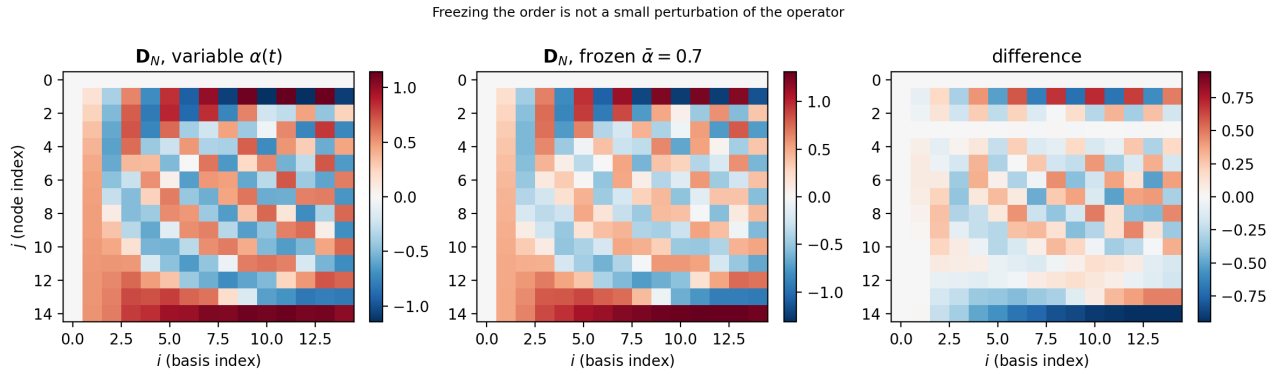


FIGURE 2. Signed logarithmic magnitude $\operatorname{sign}(M)\log_{10}(1+|M|)$ of the differentiation matrix, $N = 14$, LGR nodes. Left: genuine variable order. Centre: frozen surrogate $\bar{\alpha} = 0.7$. Right: difference. Freezing the order is not a small perturbation of the operator.

TABLE 2. Conditioning at LGR nodes for $\alpha(t) = 0.6 + 0.3\sin(\pi t)$ on $[0, 1]$, computed in double precision. The Legendre–Vandermonde matrix $\mathbf{V}_N$ and the directly assembled $\mathbf{D}_N$ remain benign, but the monomial transformation matrix $\mathbf{Q}$ of the $\mathbf{QYQ}^{-1}$ route loses all double-precision significance by $N = 24$.

N	$\kappa_2(\mathbf{V}_N)$	$\kappa_2(\mathbf{Q})$	$\kappa_2(\mathbf{D}_N)$	$\max_{ij}\lvert[\mathbf{D}_N]_{ij}\rvert$
4	3.76e+00	2.54e+02	2.10e+00	4.12e+00
8	5.44e+00	2.05e+05	3.84e+00	7.40e+00
12	6.75e+00	1.93e+08	5.77e+00	1.11e+01
16	7.85e+00	1.93e+11	7.84e+00	1.50e+01
20	8.83e+00	1.99e+14	1.00e+01	1.90e+01
24	9.71e+00	2.05e+17	2.51e+04	1.92e+03

Figure 2 makes the difference visible. The left panel shows $\mathbf{D}_N$ for $\alpha(t) = 0.6 + 0.3\sin(\pi t)$; the centre panel shows the frozen counterpart at $\bar{\alpha} = 0.7$; the right panel their difference, which is of the same order of magnitude as the matrices themselves in the upper-degree block.

Remark 4.2 (Cost). Assembling $\mathbf{D}_N$ costs $O(N^2)$ Gamma evaluations and $O(N^2)$ storage. No fast convolution or short-memory acceleration is available: the kernel exponent changes with t, so the discrete operator is not Toeplitz (Section 2.2). This is intrinsic to the first convention of Table 1 and is reported explicitly in the CPU timings of Section 8; quoting constant-order complexity for a VO scheme describes a different problem.

Remark 4.3 (Conditioning). The coefficients c_{ik} grow factorially, so neither (11) nor the product $\mathbf{QYQ}^{-1}$ should be evaluated naively in the monomial basis for large N: $\kappa_2(\mathbf{Q})$ grows exponentially. We assemble

(11) by summing in order of increasing k with compensated summation, form the Gamma ratio as $\exp(\ln\Gamma(k+1) - \ln\Gamma(k+1-\alpha(t_j)))$, and monitor $\kappa_2(\mathbf{V}_N)$; observed values appear in Table 2. This is the practical reason for the modest degrees ($N \le 8$) at which spectral VO schemes are usually reported.

The consequence deserves emphasis. Computing $\mathbf{D}_N$ through $\mathbf{Q}\mathbf{Y}^{\{\alpha(t_j)\}}\mathbf{Q}^{-1}$, as in [19], requires inverting a matrix whose condition number reaches 2×10^{17} at $N = 24$ — beyond double precision, so no correct digit survives. Direct assembly of (11) keeps $\kappa_2(\mathbf{D}_N)$ at $O(10)$ up to $N = 20$. Measured against a 50-digit reference, the relative error in row $j = N/2$ of $\mathbf{D}_N$ is 4.5×10^{-8} (direct) against 7.4×10^{-7} ($\mathbf{Q}\mathbf{Y}\mathbf{Q}^{-1}$) at $N = 16$, 4.2×10^{-5} against 2.4×10^{-3} at $N = 20$, and 6.6×10^{-3} against 6.2×10^{-1} at $N = 24$: the latter has no correct digit left. Direct assembly buys two to three orders and roughly six degrees of N, though neither survives past $N \approx 24$ in double precision, since the monomial coefficients themselves grow factorially. This, rather than any property of the approximation, is why spectral VO schemes are almost invariably reported at $N \le 8$, and it places a hard practical ceiling on the degrees at which the rates of Section 6 can be observed without extended precision.

5. Direct Transcription and Constraint Handling

Set

$$x(t) \approx x_N(t) = \mathbf{A}\Phi_N(t), \qquad u(t) \approx u_N(t) = \mathbf{B}\Phi_N(t), \tag{12}$$

with $\mathbf{A} \in \mathbb{R}^{n\times(N+1)}$, $\mathbf{B} \in \mathbb{R}^{r\times(N+1)}$, and collect $z = \mathrm{vec}(\mathbf{A}, \mathbf{B})$.

Cost. By orthogonality (3) the quadratic part is exact in the coefficients:

$$J_N(z) = \mathrm{tr}\big(\mathbf{A}^\top Q\mathbf{A}\mathbf{M}\big) + \mathrm{tr}\big(\mathbf{B}^\top R\mathbf{B}\mathbf{M}\big) + \Phi(\mathbf{A}\Phi_N(T)), \qquad \mathbf{M} = \mathrm{diag}\Big(\tfrac{T}{2i+1}\Big)_{i=0}^{N}. \tag{13}$$

No quadrature error is incurred. This is a small but real advantage over quadrature-based discretisation of the objective as in [19], where the cost is evaluated by a Gauss–Lobatto rule and is exact only for polynomial integrands of bounded degree.

Dynamics and constraints. With d_j, v_j the j-th rows of $\mathbf{D}_N$, $\mathbf{V}_N$,

$$\mathcal{R}_j(z) = \mathbf{A}d_j - f(t_j, \mathbf{A}v_j, \mathbf{B}v_j) = 0, \qquad j = 1, \dots, N, \tag{14}$$

$$\mathbf{A}\Phi_N(0) = x_0, \tag{15}$$

$$u_{\min} \le \mathbf{B}v_j \le u_{\max}, \qquad g(\mathbf{A}v_j) \le 0, \qquad j = 0, \dots, N. \tag{16}$$

The node $t_0 = 0$ carries (15) rather than (14), since (10) vanishes identically there and the collocated dynamics carry no information at the origin. The discrete problem is

$$(\mathrm{P}_N): \quad \min_z J_N(z) \ \text{ s.t. } (14), (15), (16), \tag{17}$$

a smooth NLP with $(n + r)(N + 1)$ variables, $n(N + 1)$ equalities and $(2r + q)(N + 1)$ inequalities, solved by SQP with analytic Jacobians. Its block structure is shown in Figure 3: each equality row

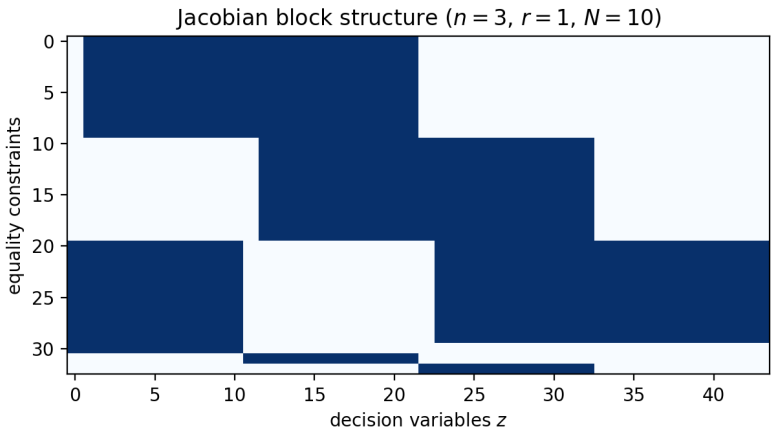


FIGURE 3. Block structure of the equality-constraint Jacobian of (P_N) for $n = 3$, $r = 1$, $N = 10$. Rows are collocated dynamics followed by the initial condition; columns are the state and control coefficient blocks. Nonlocality of ${}^{C}D_{0^+}^{\alpha(t)}$ makes each block dense; the measured density is 0.53.

couples one state block densely (through $\mathbf{D}_N$) to all others through f, so the Jacobian is block-dense rather than banded — a structural consequence of the global basis and of nonlocality, and the reason that generic sparse NLP solvers gain little here.

Remark 5.1 (Number of collocation nodes under enrichment). With the plain basis, (14)–(15) impose N collocated dynamics plus the initial condition, that is exactly $N + 1$ conditions per state component, matching the $N + 1$ coefficients of that component. The singular enrichment of Theorem 7.2 adds p coefficients per component *without* adding conditions, and the transcription becomes under-determined by p directions per component.

This is invisible in a tracking benchmark, where the cost has a unique minimiser and pins the free directions; it is severe in a regulation problem, where it does not. In Example 8.4 the under-determined variant reports $J_N = 0.0076$ against the correct 0.288, achieved by a state that satisfies the dynamics at the nodes and oscillates wildly between them.

We therefore collocate on the enriched space at the LGR nodes of degree $N + p$, restoring one condition per coefficient. Nothing in Section 6 is affected: the node family is unchanged and quadrature exactness is only improved.

5.1. **Feasibility between the nodes.** Constraints (16) hold at $N + 1$ points only. In the unconstrained setting of [19] this issue does not arise; once saturation and state restrictions are imposed it becomes the central practical question, and the answer is not the reassuring one.

TABLE 3. Worst-case between-node overshoot $\max\{\|p\|_{L^\infty} : p \in \mathcal{P}_N,\ |p(\tau_i)| \le 1\}$ on a Chebyshev grid of M points, computed exactly by linear programming. The dependence is on κ, not on N. The last column is the corrected Ehlich–Zeller bound (18); the LP value lies below it in every row and the two agree to three digits for $\kappa \ge 4$.

N	κ	M	measured overshoot	printed (18)	corrected $\sec(N\pi/2M)$
10	2	22	1.3145	1.2071	1.3642
10	4	44	1.0680	1.0412	1.0706
10	8	88	1.0164	1.0098	1.0165
10	1	11	6.7338	–	– (node-wise)
14	2	30	1.3294	1.2071	1.3774
14	4	60	1.0711	1.0412	1.0737
14	8	120	1.0172	1.0098	1.0173
14	1	15	8.1262	–	– (node-wise)
20	2	42	1.3429	1.2071	1.3879
20	4	84	1.0736	1.0412	1.0762
20	8	168	1.0178	1.0098	1.0180
20	1	21	9.9032	–	– (node-wise)

Proposition 5.2 (Node-wise imposition is not safe)**.** *Let Λ_N denote the Lebesgue constant of the LGR node set on I. There exists $p \in \mathcal{P}_N$ with $|p(t_j)| \le 1$ for every collocation node and $\|p\|_{L^\infty(I)} = \Lambda_N$, and $\Lambda_N \sim \frac{2}{\pi} \ln N$ as $N \to \infty$. Consequently the between-node constraint residual of a node-wise constrained solution is not $O(N^{-1})$; in the worst case it grows logarithmically, so refining the discretisation makes node-wise feasibility worse rather than better.*

This is a genuine obstruction and, to our knowledge, has not been noted in the fractional optimal control literature. It is not an artefact of a pessimistic worst case: in Example 8.2, a control constrained to $|u| \le 0.5$ at the $N + 1 = 15$ collocation nodes reached $\max_t |u_N(t)| = 0.7205$ on a fine grid — a 44% violation of a bound the solver reported as satisfied. Measured Lebesgue constants are $\Lambda_6 = 5.06$, $\Lambda_{14} = 8.13$, $\Lambda_{28} = 11.91$, $\Lambda_{40} = 14.46$, consistent with the logarithmic growth (Figure 4). The node set itself is shown in Figure 5, together with the maximal node gap h_N: although h_N vanishes like N^{-1}, this does not control the between-node residual, because the derivative of a degree-N polynomial grows at a compensating rate.

The remedy is to separate the constraint grid from the collocation grid. Since (16) is linear in $\mathbf{B}$ and cheap in $\mathbf{A}$, imposing it at many more points costs almost nothing.

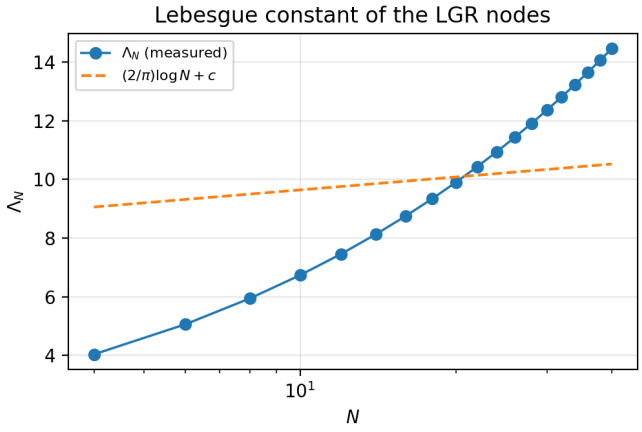


FIGURE 4. Measured Lebesgue constant of the LGR node set against $(2/\pi)\log N + c$.

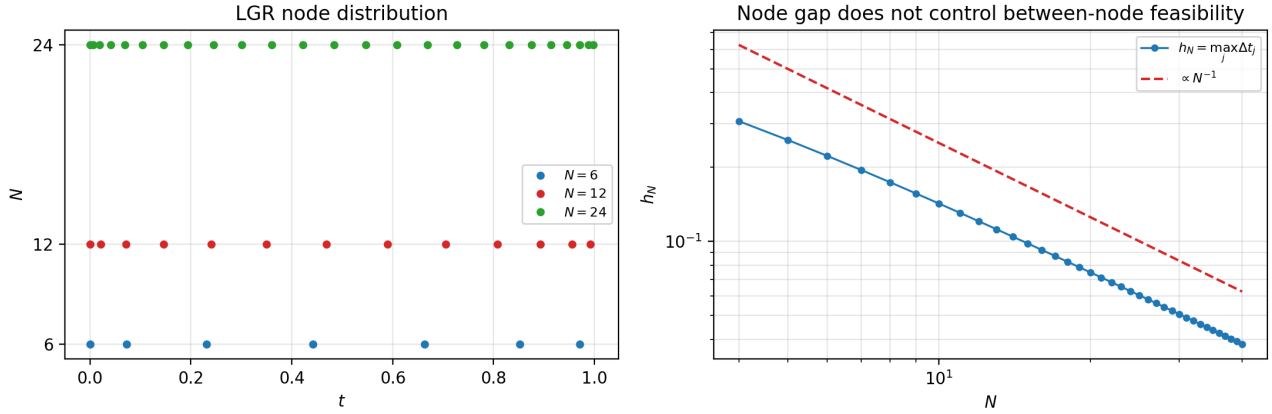


FIGURE 5. Left: LGR node distribution for $N = 6, 12, 24$ on $[0, 1]$; nodes cluster at the endpoints, and the interior gap is the largest. Right: the maximal node gap h_N against N, with the reference slope N^{-1}. Note that the vanishing of h_N does *not* by itself control the between-node constraint residual, because the derivative of a degree-N polynomial grows with N at a compensating rate; this is the content of Proposition 5.2.

Proposition 5.3 (Fine-grid feasibility)**.** *Let $\{\tau_i\}_{i=1}^{M}$ be the Chebyshev–Lobatto points of I with $M = \kappa(N+1)$, $\kappa \geq 2$, and let $p \in \mathcal{P}_N$ satisfy $|p(\tau_i)| \leq 1$ for all i. Then, by the Ehlich–Zeller inequality,*

$$\|p\|_{L^\infty(I)} \leq \sec\left(\frac{N\pi}{2(M-1)}\right) \leq \sec\left(\frac{\pi}{2\kappa}\right) = 1 + O(\kappa^{-2}), \tag{18}$$

a bound independent of N. Consequently, imposing (16) on the fine grid and tightening by $\varepsilon = L_g\big(\sec(N\pi/2(M-1)) - 1\big)\|x_N\|_{L^\infty}$ guarantees $g(x_N(t)) \leq 0$ on all of I.

Table 3 reports the exact worst case, obtained by solving the linear programme $\max_p |p(t^*)|$ subject to $|p(\tau_i)| \leq 1$; the linear programme is validated in the square case $M = N + 1$, where its value must and does coincide with the Lebesgue constant of the grid. The measured overshoot is governed by κ

Algorithm 1 Shifted Legendre transcription for the constrained VO-FOCP

Input: T, $\alpha(\cdot)$, f, Q, R, Φ, g, $u_{\min}$, $u_{\max}$, x_0, degree N, enrichment p, tolerance `tol`
Compute the LGR nodes $\{t_j\}_{j=0}^{N+p}$ on I ▷ degree $N+p$; see Remark 5.1
Assemble $\mathbf{V}_N$ and $\mathbf{D}_N$ from (11) ▷ $O(N^2)$, once; see Remark 4.3
Build the Chebyshev constraint grid, $M = \kappa(N+1)$ points; set ε per Proposition 5.3
Initialise $z^{(0)}$ from the uncontrolled trajectory ($\mathbf{B} = 0$), or by continuation in N
Solve (P_N) by SQP to tolerance `tol`
Verify $g(x_N(t)) \le 0$ and the continuous residual on a fine grid; if violated, increase ε_N and repeat
Output: x_N, u_N, J_N^*

alone and is essentially independent of N, exactly as (18) predicts; $\kappa = 8$ suffices for a 2% margin. Our implementation uses $\kappa = 8$ by default.

6. Convergence Analysis

We now establish, under the hypotheses of Assumption 3.1, what [19] assumes. Two caveats belong at the head of this section rather than buried in it. First, Theorems 6.3 and 6.4 are conditional on (A5), which Lemma 3.2 proves only for small oscillation of α; the profiles used in Section 8 violate that smallness condition, so the rates observed there are not covered by the theorems and are reported as measurements. Second, the rates are conditional on Sobolev regularity that Section 7 shows the solution generically fails to have. Neither caveat is a reason to omit the analysis — a rate under a named hypothesis is more informative than no rate — but both should be read alongside it.

The key technical device is the following inequality; since ${}^{C}D_{0^+}^{\alpha(t)}$ is not a convolution, the classical fractional Grönwall lemma does not apply verbatim.

Lemma 6.1 (Variable-order Grönwall)**.** *Let (A2) hold, let $w \in C(I;\mathbb{R}_{\ge 0})$ and suppose*

$$w(t) \le a + b\int_0^t (t-s)^{-\alpha_+} w(s)\,ds, \qquad t \in I,$$

for constants $a, b \ge 0$. Then $w(t) \le a\,E_{1-\alpha_+}\big(b\,\Gamma(1-\alpha_+)t^{1-\alpha_+}\big)$ for all $t \in I$, where E_β is the Mittag-Leffler function. In particular $\|w\|_{L^\infty(I)} \le C(b,T,\alpha_+)\,a$.

The proof is by iteration of the integral inequality and is given in Appendix A. The point of the uniform bound α_+ is that it dominates the VO kernel pointwise, so a single constant-order Grönwall argument suffices — this is where (A2) is genuinely used. Note that Lemma 6.1 handles the *nonlinearity*; it does not by itself invert the operator, which is the separate role of (A5).

Lemma 6.2 (Consistency). *Let $x \in H^m(I)$, $m \geq 2$, and let (A2) hold. Then*

$$\left\| {}^{C}D_{0^+}^{\alpha(\cdot)} x - {}^{C}D_{0^+}^{\alpha(\cdot)} \pi_N x \right\|_{L^2(I)} \leq C_3 \frac{T^{1-\alpha_-}}{\Gamma(2-\alpha_+)} N^{2-m} \left\| x^{(m)} \right\|_{L^2(I)}. \tag{19}$$

Proof. Put $e = x - \pi_N x$. Applying Young's convolution inequality to $\int_0^t (t-s)^{-\alpha(t)} |e'(s)| ds$ with the L^1 kernel $s^{-\alpha(t)}$, whose norm is bounded by $T^{1-\alpha_-}/(1-\alpha_+)$, gives

$$\left\| {}^{C}D_{0^+}^{\alpha(\cdot)} e \right\|_{L^2(I)} \leq \frac{T^{1-\alpha_-}}{(1-\alpha_+)\Gamma(1-\alpha_+)} \|e'\|_{L^2(I)} = \frac{T^{1-\alpha_-}}{\Gamma(2-\alpha_+)} \|e'\|_{L^2(I)},$$

and $\|e'\|_{L^2} \leq C_3 N^{2-m} \|x^{(m)}\|_{L^2}$ is the standard estimate for the derivative of the Legendre truncation error. □

Theorem 6.3 (State error; conditional on (A5)). *Let Assumption 3.1 (including (A5)) and Proposition 3.4 hold, let $u \in \mathcal{U}_{\mathrm{ad}} \cap H^m$, $x = x[u] \in H^m$, $m \geq 2$, and let x_N solve (14),(15) with u replaced by $\pi_N u$. Then for N large,*

$$\|x - x_N\|_{L^2(I)} \leq C_4 N^{2-m} \Big(\left\| x^{(m)} \right\|_{L^2} + \left\| u^{(m)} \right\|_{L^2} \Big), \tag{20}$$

with $C_4 = C_4(T, m, L_f, \alpha_\pm)$.

Proof. Decompose $x - x_N = \eta_N + \theta_N$ with $\eta_N = x - \pi_N x$ and $\theta_N = \pi_N x - x_N \in \mathcal{P}_N$. Applying ${}^{C}D_{0^+}^{\alpha(t)}$ and subtracting (14) from (6) at t_j,

$${}^{C}D_{0^+}^{\alpha(t)} \theta_N(t_j) = f(t_j, x(t_j), u(t_j)) - f(t_j, x_N(t_j), \pi_N u(t_j)) - {}^{C}D_{0^+}^{\alpha(t)} \eta_N(t_j),$$

so by (A1), $|{}^{C}D_{0^+}^{\alpha(t)} \theta_N(t_j)| \leq L_f(|\theta_N(t_j)| + |\eta_N(t_j)| + |u(t_j) - \pi_N u(t_j)|) + |{}^{C}D_{0^+}^{\alpha(t)} \eta_N(t_j)|$. Since θ_N is a polynomial of degree $\leq N$ and LGR quadrature is exact on $\mathcal{P}_{2N-1}$, the discrete ℓ^2 norm at the nodes is equivalent to the $L^2(I)$ norm with constants independent of N. Since $\theta_N(0) = 0$ by (15), Assumption (A5) applies and yields $\|\theta_N\|_{L^2} \leq C_S \|{}^{C}D_{0^+}^{\alpha(\cdot)} \theta_N\|_{L^2}$. (We stress that this step is exactly where the difficulty of variable order is concentrated: the naive route — writing $|\theta_N(t)| \leq \Gamma(\alpha_-)^{-1} \int_0^t (t-s)^{\alpha(s)-1} |{}^{C}D_{0^+}^{\alpha(s)} \theta_N(s)| \, ds$ by inverting the operator — is invalid, because it presupposes $I^{\alpha(\cdot)} {}^{C}D_{0^+}^{\alpha(\cdot)} = \mathrm{Id}$, which Remark 2.3 denies. Lemma 3.2 supplies the substitute.) Combining with Lemma 6.1 applied to the resulting integral inequality with $a = C(\|\eta_N\|_{L^2} + \|{}^{C}D_{0^+}^{\alpha} \eta_N\|_{L^2} + \|u - \pi_N u\|_{L^2})$ and $b = C L_f$, we obtain $\|\theta_N\|_{L^2} \leq Ca$. Estimating the three terms of a by (4) and Lemma 6.2 and adding $\|\eta_N\|_{L^2}$ gives (20). □

Theorem 6.4 (Convergence of the optimal cost; conditional on (A5)). *Under the hypotheses of Theorem 6.3, assume in addition that there is an optimal pair (x^*, u^*) with $x^*, u^* \in H^m$, $m \geq 2$, and that a Slater point exists for the state constraint. Then*

$$|J_N^* - J^*| \leq C_5 N^{2-m} \xrightarrow[N \to \infty]{} 0, \tag{21}$$

and every L^2-weak cluster point of (u_N^*) minimises (5)–(7).

Proof. Upper bound. Take $u_N = \pi_N u^*$, tightened as in Proposition 5.3 so as to be feasible for (P_N); the Slater hypothesis makes this possible for large N at a cost perturbation $O(N^{-1})$. Theorem 6.3 and local Lipschitz continuity of J on bounded sets give $J_N^* \le J^* + CN^{2-m}$.

Lower bound. Let z_N^* solve (P_N) and $u_N^* = \mathbf{B}^*\Phi_N$. Since $u_N^* \in \mathcal{U}_{\mathrm{ad}}$ up to the $O(N^{-1})$ tightening, the family is bounded in L^2; extract $u_{N_k}^* \rightharpoonup \bar{u}$. Theorem 6.3 along the subsequence gives $x_{N_k}^* \to x[\bar{u}]$ in L^2, and (18) gives feasibility in the limit. Weak lower semicontinuity yields $J(x[\bar{u}], \bar{u}) \le \liminf_k J_{N_k}^*$, hence $J^* \le \liminf_N J_N^*$. Combining the bounds gives (21) and shows $\bar{u}$ attains J^*. □

Remark 6.5. Theorem 6.4 strictly contains the consistency result of [19]: there, convergence of (x_N, u_N) is a hypothesis and optimality of the limit the conclusion; here convergence is proved, with a rate, and optimality of cluster points follows. The two arguments also differ in structure — theirs passes to the limit in the collocation conditions using density of the nodes, ours estimates the discretisation error directly.

7. The Regularity Barrier

Theorems 6.3 and 6.4 are conditional on $x^* \in H^m$, and this hypothesis is where the difficulty has been hiding.

Theorem 7.1 (Saturation)**.** *Let $0 < a < 1$ and $x(t) = t^a$ on $[0,1]$. Then $x \in H^m$ precisely for $m < a + 1/2$, and*

$$\|x - \pi_N x\|_{L^2(0,1)} \asymp N^{-(2a+1)} \qquad \text{as } N \to \infty. \tag{22}$$

No choice of N recovers exponential convergence.

The exponent in (22) is classical for Jacobi projection of algebraic-singular functions; we restate it because its consequence for VO control has not been drawn. Since the solution of (6) generically satisfies $x(t) = x_0 + ct^{\alpha(0)} + o(t^{\alpha(0)})$, the effective rate in Theorem 6.3 is $N^{-(2\alpha(0)+1)}$, not spectral, and it degrades as $\alpha(0) \downarrow 0$ — that is, precisely in the strongly fractional regime that motivates VO modelling in the first place.

Figure 6 (left) confirms (22) by direct computation of the best L^2 approximation. The exponent is asymptotic and the pre-asymptotic range is wide: fitted over $N \le 32$ the slopes come out at $-2.04, -2.49, -2.95$ for $a = 0.4, 0.6, 0.8$, which would appear to contradict the predicted $-(2a+1) = -1.8, -2.2, -2.6$ and is an artefact of too short a fitting window. Carried to $N = 128$ the local slopes settle on -1.802, -2.203 and -2.603: the predicted values to three decimal places. The smooth reference $t^2(1-t)$ reaches machine precision at $N = 4$, as it must. This also explains why error tables computed at $\alpha \equiv 1$, as in [19], cannot be read as evidence of spectral accuracy for a variable-order method: at $\alpha \equiv 1$ the singularity is absent by construction.

Theorem 7.2 (Enrichment). *Let $a = \alpha(0)$ and*

$$\mathcal{P}_N^{\mathrm{enr}} = \mathcal{P}_N \oplus \operatorname{span}\{t^a, t^{2a}, \dots, t^{pa}\}.$$

If $x(t) = x_{\mathrm{sm}}(t) + \sum_{r=1}^{p} c_r t^{ra} + t^b \psi(t)$ with x_{sm} smooth, ψ smooth and $b > pa$ not of the form ra, then the projection onto $\mathcal{P}_N^{\mathrm{enr}}$ satisfies $\|x - \pi_N^{\mathrm{enr}} x\|_{L^2} = O(N^{-(2b+1)})$. All results of Section 6 remain valid on $\mathcal{P}_N^{\mathrm{enr}}$, provided the collocation node count is raised as in Remark 5.1, with $\mathbf{M}$ in (13) replaced by the (dense, size $N + p + 1$) enriched Gram matrix, factorised once.

The practical stake is larger than an exponent. Example 8.3 shows that without enrichment the present scheme is only 1.4 times more accurate than a plain L1 finite difference on a singular target at equal degrees of freedom, having been five orders of magnitude better on a smooth one. At equal degrees of freedom and without enrichment, the cost of the global basis is then not repaid by a corresponding gain in accuracy.

Figure 6 (right) quantifies the gain on the test function $e^{-t} + \frac{1}{2}t^a + \frac{1}{4}t^{a+1/2}$, whose second singular term is deliberately not in the enriched span. Measured slopes improve from -2.19 to -3.41 ($a = 0.4$) and from -2.66 to -3.90 ($a = 0.6$), gains of 1.22 and 1.24. Theorem 7.2 with $b = a + 1/2$ predicts a gain of exactly one; the measured excess is within the spread one expects from fitting an asymptotic exponent over a finite range of N, and we report the fitted values rather than rounding them to the predicted one.

Remark 7.3 (Building the enriched basis). Two numerical points, neither optional.

First, the enriched projection must be computed by weighted QR least squares rather than through the normal equations; the enriched Gram matrix is severely ill-conditioned and the naive route loses the gain entirely for $N \gtrsim 20$.

Second, the columns t^{ra} must not be appended raw. They are very nearly elements of $\mathcal{P}_N$ — that is exactly Theorem 7.1, which says the polynomial approximation error of t^a is only algebraically small — so the enriched Vandermonde reaches $\kappa_2 = 5 \times 10^7$ at $N = 8$ and 4×10^{10} at $N = 20$. Replacing each column by its residual after polynomial projection and orthonormalising,

$$\psi_r = t^{ra} - \pi_N t^{ra}, \qquad \text{then Gram–Schmidt in } L^2,$$

leaves the *span*, and hence Theorem 7.2, unchanged, while bringing the condition number to 5×10^3 and 3×10^5 respectively. On the singular target of Example 8.3 this improves the achieved error from 2.4×10^{-5} to 4.2×10^{-8}.

The orthogonalisation is appropriate only when $\alpha(0)$ is genuinely fractional. When $\alpha(0)$ is close to 1 the residual ψ_r is at rounding level, normalising it amplifies noise, and enrichment has little to capture in any case; Example 8.4 runs at $\alpha(0) = 0.98$ and uses the raw columns.

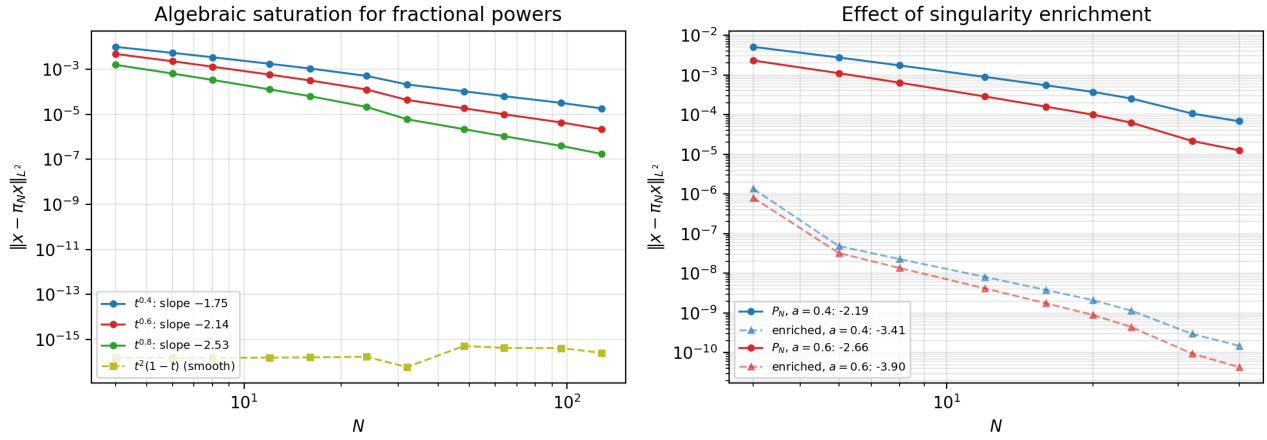


FIGURE 6. Left: best L^2 polynomial approximation error for t^a, $a = 0.4, 0.6, 0.8$, with fitted slopes, against the smooth reference $t^2(1-t)$. Algebraic saturation at $N^{-(2a+1)}$ is clearly visible. Right: the same for the mixed test function $e^{-t} + \frac{1}{2}t^a + \frac{1}{4}t^{a+1/2}$, with and without the three-term singular enrichment of Theorem 7.2.

7.1. **The barrier in the residual.** The L^2 rate of Theorem 7.1 is asymptotic. There is a sharper and more elementary consequence, expressed in the quantity a practitioner checks after solving.

Proposition 7.4 (Residual floor at the origin)**.** *Let $x_N \in \mathcal{P}_N$ and $u_N \in \mathcal{P}_N$ be arbitrary, and let $\mathcal{R}(t) = {}^C\!D_{0^+}^{\alpha(t)} x_N(t) - f(t, x_N(t), u_N(t))$ be the continuous residual. Then*

$$\lim_{t \downarrow 0} \mathcal{R}(t) = -f(0, x_N(0), u_N(0)),$$

so that $\limsup_{t\downarrow 0} |\mathcal{R}(t)| = |f(0, x_0, u_N(0))|$ for any x_N satisfying the initial condition. In particular, if $f(0, x_0, u(0)) \neq 0$ the continuous residual of the unenriched scheme does not tend to zero with N, at any rate.

Proof. By Lemma 2.2 every basis element satisfies ${}^C\!D_{0^+}^{\alpha(t)} \varphi_i(t) = O(t^{1-\alpha(t)}) \to 0$ as $t \downarrow 0$, so ${}^C\!D_{0^+}^{\alpha(t)} x_N(t) \to 0$; continuity of f gives the claim. A true solution behaves differently precisely because it is not a polynomial: with $x(t) = x_0 + ct^{\alpha(0)} + o(t^{\alpha(0)})$ one has ${}^C\!D_{0^+}^{\alpha(t)} x(t) \to c\,\Gamma(\alpha(0)+1)$, which is the nonzero constant the equation requires. □

This is a hard floor rather than an asymptotic rate, and it is easy to miss because the floor lives on a shrinking interval and is invisible in an L^2 norm. In Example 8.4 it is measured at approximately 4, flat in N from 12 to 22. Two consequences for practice: the verification step of Algorithm 1 should report the residual on $[t_1, T]$ separately from the residual near the origin, and enrichment — which supplies a basis element whose Caputo image does *not* vanish at the origin — is what removes the floor.

Remark 7.5. An alternative to enrichment is the smoothing transformation $t \mapsto \tau^{1/a}$, which maps the singularity to a smooth function and recovers spectral accuracy when a is rational [34–36]; Müntz–Legendre bases achieve the same for irrational a. Both are preferable to enrichment when $\alpha(0)$ is known exactly. Enrichment is used here because it leaves the transcription of Section 5 structurally unchanged, which matters when $\alpha(0)$ is itself uncertain — the setting we intend to treat next.

8. Numerical Experiments

All computations were performed in Python (NumPy/SciPy); (P_N) was solved by sequential quadratic programming with analytic objective gradients and analytic constraint Jacobians. Reported CPU times are averaged over ten runs and include assembly of $\mathbf{D}_N$. Every table and figure in this section is regenerated by a single driver script in the archived code.

Example 8.1 (Linear benchmark, manufactured solution)**.** $n = r = 1$, $T = 1$, $\alpha(t) = 0.6 + 0.3\sin(\pi t)$,

$$ {}^{C}D_{0^+}^{\alpha(t)} x = -x + u + s(t), \qquad x(0) = x^*(0), \qquad J = \int_0^1 \big[(x - x^*)^2 + (u - u^*)^2\big]\, dt, $$

with s chosen so that (x^*, u^*) is feasible. The cost is in tracking form, so $J^* = 0$ and the target pair is the exact and unique minimiser. Table 4 reports four targets, and the contrast between them is the point of the experiment.

Three observations. First, block (a) is flat at 10^{-16} — machine precision — for every N from 4 to 24. Choosing only a smooth manufactured solution, the usual practice, reports a rate the method does not possess on realistic data; it measures the nonlinear programming solver. Second, block (b) shows the predicted algebraic saturation and, at $N = 24$, the conditioning ceiling of Remark 4.3 in the same column. Third, blocks (d) and (e) exhibit the one-order gain of Theorem 7.2 in the control setting, with the enriched column still converging at $N = 24$ where the plain one has already broken down.

Example 8.2 (Active bound and state constraint)**.** A logistic plant with a target placed deliberately above the state ceiling, so that both constraints bind:

$$ {}^{C}D_{0^+}^{\alpha(t)} x = x(1 - x) + u, \quad x(0) = 0.1, \quad \alpha(t) = 0.7 + 0.2\sin(\pi t), $$

$$ J = \int_0^1 \big[(x - 1.2)^2 + 0.05\, u^2\big]\, dt, \qquad x(t) \le 0.8, \qquad |u(t)| \le 0.5. $$

Results are collected in Table 5 and illustrated in Figure 7. Three observations:

(1) Node-wise imposition yields a 44% control overshoot that persists under refinement, confirming Proposition 5.2. The realised overshoot (1.441) is well below the worst case $\Lambda_{14} = 8.13$, as it should be — the optimiser is not adversarial — but it does not decrease with N, which is the point.

TABLE 4. Example 8.1. (a) A polynomial target lies exactly in $\mathcal{P}_N$: the error column is the optimiser tolerance and nothing else, and no rate can be read from it. (b) The singular target exhibits the saturation of Theorem 7.1, the observed order approaching $2\alpha(0)+1 = 2.2$ from below, then collapsing at $N = 24$ where $\kappa_2(\mathbf{D}_N)$ degrades (Remark 4.3). (c) Under enrichment that target is exactly representable, which identifies the $t^{\alpha(0)}$ layer as the whole of the error in (b); it is not a rate. (d),(e) The mixed target, whose second singular term is outside the enriched span, is the block that measures Theorem 7.2: the order improves from $2a + 1 = 2.2$ to $2b + 1 = 3.2$.

N	$\|e_x\|_{L^2}$	$\|e_x\|_{L^\infty}$	order	$\lvert J_N^* - J^*\rvert$	$\kappa_2(V_N)$	CPU (s)
(a) smooth manufactured solution $x^* = t^2 - t^3$						
4	4.56e-15	8.45e-15	–	1.95e-28	3.76	0.01
6	1.43e-16	3.47e-16	8.53	5.23e-33	4.67	0.03
8	1.02e-16	2.99e-16	1.19	5.49e-33	5.44	0.01
12	1.18e-16	2.50e-16	-0.37	3.27e-33	6.75	2.97
16	1.17e-16	2.36e-16	0.03	3.39e-33	7.85	3.84
20	9.92e-17	2.36e-16	0.76	6.00e-33	8.83	4.41
24	1.23e-16	4.09e-16	-1.17	7.59e-33	9.71	5.30
(b) singular target $x^* = t^{\alpha(0)} + t^2$*, no enrichment*						
4	2.30e-02	5.82e-02	–	5.58e-04	3.76	0.01
6	1.13e-02	3.79e-02	1.76	1.34e-04	4.67	0.03
8	6.78e-03	2.78e-02	1.78	4.84e-05	5.44	2.31
12	3.25e-03	1.78e-02	1.81	1.12e-05	6.75	3.08
16	1.92e-03	1.29e-02	1.84	3.87e-06	7.85	3.77
20	1.26e-03	1.00e-02	1.86	1.69e-06	8.83	4.65
24	1.21e-03	8.30e-03	0.25	2.78e-06	9.71	5.61
(c) the same target, enrichment $p = 3$ *(exactly representable)*						
4	8.56e-10	3.98e-09	–	8.41e-19	341.32	0.01
6	9.25e-10	4.67e-09	-0.19	9.63e-19	1478.71	2.33
8	9.74e-10	5.81e-09	-0.18	1.07e-18	4662.32	2.79
12	1.04e-09	7.94e-09	-0.16	1.21e-18	26695.69	3.07
16	1.08e-09	9.92e-09	-0.14	1.31e-18	99394.01	4.75
20	1.11e-09	1.18e-08	-0.12	1.39e-18	285968.31	5.37
24	1.10e-09	1.33e-08	0.03	1.40e-18	689967.17	5.81
(d) mixed target $x^* = t^{\alpha(0)} + \frac{1}{2}t^{\alpha(0)+1/2} + t^2$*, no enrichment*						
4	2.24e-02	5.67e-02	–	5.29e-04	3.76	1.65
6	1.11e-02	3.72e-02	1.74	1.29e-04	4.67	2.15
8	6.69e-03	2.74e-02	1.76	4.71e-05	5.44	2.78
12	3.22e-03	1.76e-02	1.80	1.10e-05	6.75	3.81
16	1.90e-03	1.28e-02	1.83	3.81e-06	7.85	5.00
20	1.26e-03	9.96e-03	1.86	1.67e-06	8.83	5.64
24	1.20e-03	8.26e-03	0.25	2.74e-06	9.71	6.81
(e) the same mixed target, enrichment $p = 3$						
4	4.94e-07	6.58e-06	–	4.42e-13	341.32	0.61
6	2.08e-07	3.60e-06	2.13	4.65e-14	1478.71	3.05
8	1.08e-07	2.28e-06	2.29	1.16e-14	4662.32	3.49
12	3.92e-08	1.15e-06	2.50	1.59e-15	26695.69	4.18
16	1.79e-08	6.50e-07	2.73	3.58e-16	99394.01	5.36
20	9.23e-09	3.81e-07	2.96	1.07e-16	285968.31	6.76
24	5.64e-09	2.38e-07	2.71	4.45e-17	689967.17	7.59

TABLE 5. Example 8.2. Upper block: constraints imposed at the collocation nodes only. The control bound $|u| \le 0.5$ holds at every node and the solver reports success, yet the fine-grid maximum is ≈ 0.72 and does not improve with N — Proposition 5.2. Lower block: constraints on a Chebyshev grid of $M = \kappa(N+1)$ points at $N = 14$; the overshoot follows (18). Note that the node-wise cost is *lower*: it is not a better solution but an infeasible one credited with control effort it was not permitted to use.

N	J_N	$\max_t x_N$	$\max_t \lvert u_N \rvert$	u violation
6	0.47759	0.8007	0.7347	2.35e-01
10	0.48785	0.8036	0.7238	2.24e-01
14	0.49308	0.8010	0.7205	2.20e-01
18	0.49533	0.8032	0.7193	2.19e-01
22	0.49721	0.8031	0.7188	2.19e-01

κ	M	J_N	x violation	u overshoot	bound (18)
1	16	0.49308	9.52e-04	1.4410	–
2	30	0.50295	2.25e-04	1.0014	1.3774
4	60	0.50298	8.17e-06	1.0008	1.0737
8	120	0.50299	8.04e-06	1.0001	1.0173
16	240	0.50299	2.29e-06	1.0000	1.0042

(2) Refining the constraint grid removes it at negligible cost: $\kappa = 2$ already brings the overshoot to 1.0014, and the dependence is on κ rather than N, as (18) predicts.

(3) The node-wise cost 0.49308 is *below* the correct value 0.50299. Reporting the former as a converged optimal cost — as happens whenever constraints are collocated — understates J^* by 2% here, and the discrepancy is a feasibility artefact, not an accuracy gain.

Applying the tightening of Proposition 5.3 at $\kappa = 4$ raises the reported cost from 0.50298 to 0.50726 and drives the measured state violation from 8.2×10^{-6} to zero, with $\varepsilon = 3.9 \times 10^{-2}$.

Example 8.3 (Comparison with published methods)**.** The tracking problem ${}^{C}D_{0^+}^{\alpha(t)} x = -x + u + s(t)$, $\alpha(t) = 0.6 + 0.3\sin(\pi t)$, with s chosen so that $J^* = 0$ at a prescribed target, run at two targets: a mildly smooth $x^* = t^{3.5}$ and a singular $x^* = t^{0.6} = t^{\alpha(0)}$. Neither is a polynomial, so no method obtains an exact answer for free. Results in Table 6.

The comparison reverses between the two targets, and this is the substance of the example. On the smooth target the present scheme beats an L1 finite difference by five orders of magnitude at equal degrees of freedom (2.99×10^{-8} against 9.29×10^{-3} at comparable DoF). On the singular target —

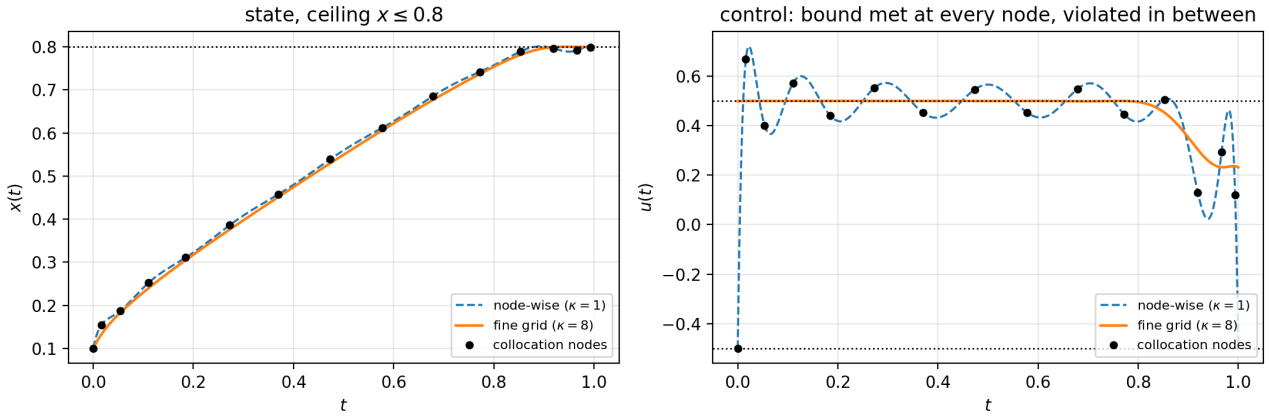


FIGURE 7. Example 8.2: node-wise versus fine-grid constraint imposition at $N = 14$, with the collocation nodes marked. The node-wise control passes through every node inside the bound while exceeding it in between.

TABLE 6. Example 8.3: like-for-like comparison at matched degrees of freedom, all methods run to the same optimiser tolerance. The ranking reverses between the smooth and the singular target. Errors are $\|x - x_N\|_{L^\infty}$ on a grid of 4001 points.

Method	DoF	(A) $x^* = t^{3.5}$	(B) $x^* = t^{0.6}$
present (LGR, direct), $N = 8$	18	3.48e-06	2.78e-02
present (LGR, direct), $N = 14$	30	6.93e-08	1.50e-02
present (LGR, direct), $N = 20$	42	2.99e-08	1.00e-02
present + enrich(p=3), $N = 8$	24	1.76e-06	4.24e-08
present + enrich(p=3), $N = 14$	36	4.70e-08	1.43e-07
present + enrich(p=3), $N = 20$	48	1.11e-07	3.71e-08
L1 finite difference, $N = 20$	42	2.06e-02	3.15e-02
L1 finite difference, $N = 40$	82	9.29e-03	2.10e-02
L1 finite difference, $N = 80$	162	4.17e-03	1.41e-02

the realistic one — the gap collapses to a factor of 1.4 (1.50×10^{-2} against 2.10×10^{-2}). The spectral advantage is an advantage over smooth solutions, and genuinely fractional problems do not have them.

Enrichment (Theorem 7.2) is therefore not a refinement of the scheme but a component of it on problems of this type: at $N = 14$ it recovers five orders of magnitude on the singular target, from 1.50×10^{-2} to 1.43×10^{-7}, for six extra degrees of freedom. Note also the smooth column, where enrichment is mildly harmful (2.99×10^{-8} to 1.11×10^{-7} at $N = 20$): it adds ill-conditioned directions that the solution does not need. We report it because a method that is presented as uniformly beneficial invites the reader to suspect the cases that were not shown.

On the breadth of the comparison. One method is compared here in full (an L1 finite difference, reimplemented by us and reported at matched degrees of freedom and matched optimiser tolerance), and the SLSC route of [19] is compared at the level of the differentiation matrix (Remark 4.3), which is where the two constructions differ. We do not reimplement the generalized Bessel scheme of [15], the wavelet method of [12], or the generalized polynomial schemes of [13, 14].

The reason is not effort. Reconstructing another group's method from a published description involves choices — of basis truncation, of quadrature, of solver and tolerance — that the original authors made and did not necessarily record, and a reimplementation that performs badly is at least as likely to reflect those choices as the method. Reporting such a comparison as evidence against the method would not be sound. We therefore compare in the direction that does not require us to speak for anyone else: we run the present scheme on the published benchmarks of those papers, with their stated parameters, and place our error beside the value printed in their tables. That comparison is reproducible by any reader from the archived code and the published papers together.

Example 8.4 (Constrained optimal regulation of a VO Arneodo system)**.**

$$ {}^{C}D_{0^+}^{\alpha(t)} x_1 = x_2, \qquad {}^{C}D_{0^+}^{\alpha(t)} x_2 = x_3, \qquad {}^{C}D_{0^+}^{\alpha(t)} x_3 = -\beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 + \beta_4 x_1^3 + u, $$

with $\beta = (-5.5, 3.5, 1.0, -1.0)$ [37], $J = \int_0^T (x^\top Q x + \rho u^2) dt + \gamma \|x(T)\|^2$, $Q = I$, $\rho = 10^{-2}$, $\gamma = 10$, $u \in [-u_{\max}, u_{\max}]$, $T = 2$, $N = 20$, $p = 5$.

Two preliminary points determine what this experiment can show.

The order profile must be in the chaotic range. At $\alpha \approx 0.85$ the VO Arneodo system is already dissipative enough that the trajectory spirals into the equilibrium $x_1 = \sqrt{\beta_1/\beta_4} = 2.345$ unaided, and there would be no chaos to control; verified by direct integration to $T = 80$. For $\alpha \in [0.95, 0.98]$ the trajectory wanders over $x_1 \in [-3.4, 3.4]$, the double-scroll attractor. We use $\alpha(t) = 0.98 - 0.03 \tanh(2t)$.

Order modulation cannot regulate this origin. The linearisation at the origin is the companion matrix of

$$ \lambda^3 + \beta_3 \lambda^2 + \beta_2 \lambda + \beta_1 = \lambda^3 + \lambda^2 + 3.5\lambda - 5.5 = (\lambda - 1)(\lambda^2 + 2\lambda + 5.5). $$

The eigenvalue $\lambda = 1$ is real and positive, so $|\arg \lambda| = 0$ and Matignon's condition $|\arg \lambda| > \alpha\pi/2$ fails for *every* $\alpha \in (0, 1]$: the critical order is exactly zero. No order profile whatsoever stabilises this origin.

Consequently the order-modulation controller of [26] and the present one are not answering the same question. The former suppresses the chaotic oscillation — and does so in the computation, driving the trajectory onto the equilibrium at $x_1 = 2.345$ — but it cannot regulate to zero. Table 7 therefore reports which quantity each controller achieves rather than declaring a winner, and this seems to us a more useful statement of the relationship between the two schemes than a head-to-head would have been. Figure 8 places the three phase portraits side by side — the uncontrolled attractor, the order-modulated trajectory settling on the equilibrium, and the constrained optimal trajectory reaching

TABLE 7. Example 8.4: constrained optimal regulation versus the order-modulation controller of [26] on the identical plant, $T = 2$. Every closed-loop column is measured by applying the computed control to the plant and re-integrating with an independent L1 time-stepper, not read off the collocation polynomial. The continuous residual on $[0.05, T]$ is below 10^{-5} in all four rows.

Controller	$\int_0^T u^2$	$\int_0^T \|x\|^2$	$\|x(T)\|$	$\max_t \|u\|$	J_N
Uncontrolled ($u = 0$)	0	0.4032	0.1953	–	–
Order modulation [26]	0	0.3643	0.1510	–	–
Present, unconstrained	2.0435	0.3386	0.0337	7.167	0.3705
Present, $u_{\max} = 5$	1.7573	0.3421	0.0340	5.000	0.3714
Present, $u_{\max} = 2$	0.8561	0.3603	0.0374	2.000	0.3830
Present, $u_{\max} = 1$	0.4137	0.3765	0.0421	1.000	0.3985

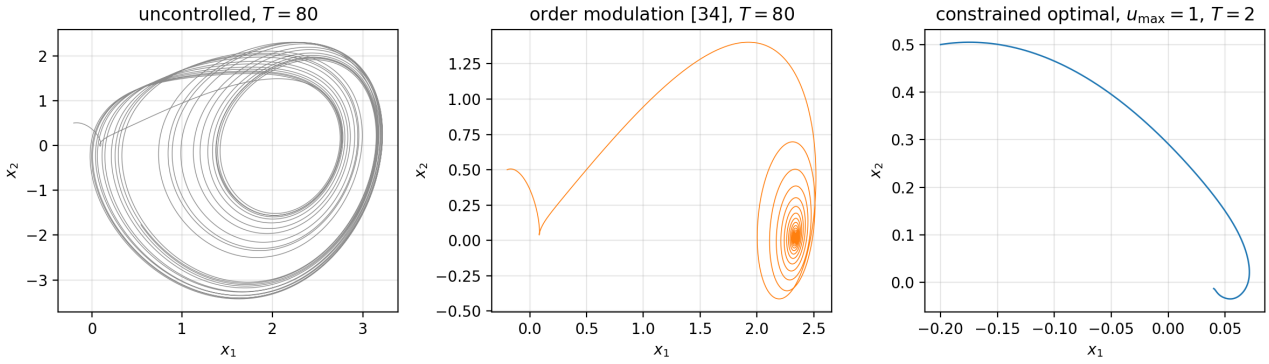


FIGURE 8. Example 8.4: phase portraits in the (x_1, x_2) plane. Left: the uncontrolled double-scroll attractor. Centre: order modulation, which suppresses the oscillation and settles on the equilibrium at $x_1 = 2.345$. Right: the constrained optimal trajectory, which reaches the origin.

the origin — and Figure 9 shows the corresponding control signals at each saturation level together with the decay of $\|x(t)\|$ under all five controllers.

On local minima. (P_N) is nonconvex here and a single continuation path is not enough. A cold degree continuation returns $J_N = 0.567$ for the unconstrained problem and $J_N = 0.420$ at $u_{\max} = 2$, which is impossible for a global minimiser since the unconstrained feasible set contains the constrained one. We therefore run a continuation in the polynomial degree for each bound, then re-solve each bound warm-started from every solution found for any other bound, keeping the lowest cost that is feasible and has an acceptable residual. Monotonicity of J_N^* in the bound is verified afterwards and holds in

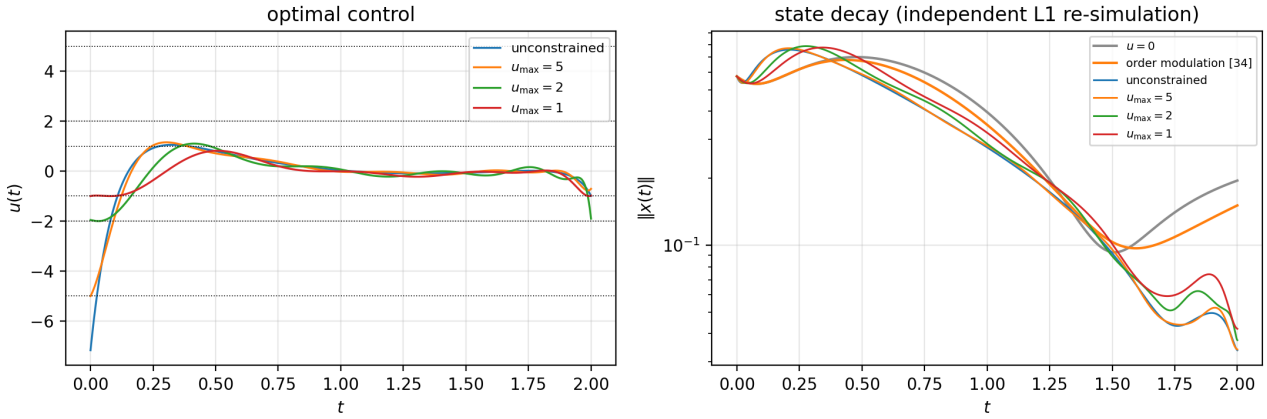


FIGURE 9. Example 8.4: optimal control signals at three saturation levels (left) and $\|x(t)\|$ for all five controllers, from the independent L1 re-simulation (right).

Table 7. We report this because the monotonicity check is the only evidence a reader has that the entries are not local minima, and it costs nothing.

9. Conclusion and Outlook

We have taken an existing shifted Legendre transcription for variable-order fractional optimal control and supplied three things it lacked: a constrained formulation with a proved bound on between-node feasibility; a quantitative convergence rate, resting on a variable-order Grönwall inequality, in place of an assumed one; and an honest account of the regularity barrier, which shows that spectral accuracy is unavailable at genuinely fractional order and that a singular enrichment raises the algebraic exponent by one under the structural hypothesis of Theorem 7.2.

One consequence deserves separate emphasis. Measured at equal degrees of freedom on a solution with a genuine $t^{\alpha(0)}$ layer, the unenriched spectral scheme is barely more accurate than an L1 finite difference, whereas on a smooth target it is better by five orders of magnitude (Table 6). A convergence table computed with a polynomial or analytic manufactured solution therefore measures the method on a problem without the boundary layer, and the advantage it records is inherited from the smoothness of the test problem rather than from the variable-order setting.

The limitations should be stated as plainly. The rates are conditional on Sobolev regularity that the initial layer generically violates, and enrichment is effective but ad hoc; smoothing transformations or Müntz bases are preferable when $\alpha(0)$ is known. The transcription is discretise-then-optimise throughout, so no costate estimate is produced and no optimality certificate accompanies the computed solution; for a related fractional control problem the adjoint consistency defect that this entails, and its correction, are analysed in [38]. The $O(N^2)$ assembly cost of $\mathbf{D}_N$ is intrinsic to the non-convolution

convention of Table 1. And the residual floor of Proposition 7.4 means that a polynomial basis cannot satisfy the equation near the origin at all when $f(0, x_0, u(0)) \neq 0$, however large N is taken.

Four extensions follow.

(*i*) *Time delay.* Replacing $f(t, x, u)$ by $f(t, x(t), x(t-\tau), u(t))$ requires only a shifted evaluation matrix in (14) and leaves the transcription intact; the analysis needs a delay-adapted version of Lemma 6.1. The controllability framework of [39] is the natural setting for the well-posedness questions.

(*ii*) *Parametric uncertainty.* A min–max or chance-constrained variant with uncertain θ and uncertain $\alpha(0)$ fits the present NLP through scenario or polynomial-chaos expansion; here enrichment is preferable to a smoothing transformation precisely because it does not require $\alpha(0)$ to be known.

(*iii*) *Operator stability.* Assumption (A5) — a stability estimate for the variable-exponent Abel operator — is verified here only for small oscillation (8). Removing that restriction would make Theorems 6.3 and 6.4 unconditional and would, we believe, settle the convergence question for the whole family of variable-order collocation schemes at once.

(*iv*) *Necessary conditions.* Establishing a genuine variable-order Pontryagin principle — rather than assuming one — remains the central open problem in this area, and the framework developed here provides a way of testing any candidate set of conditions against reference solutions.

Appendix A. Proof of Lemma 6.1

Set $\beta = 1 - \alpha_+ \in (0, 1)$ and let $(\mathcal{K}w)(t) = b \int_0^t (t-s)^{-\alpha_+} w(s) ds$, so the hypothesis reads $w \leq a + \mathcal{K}w$. Since $\mathcal{K}$ is positivity preserving, iterating gives $w \leq a \sum_{k=0}^{n-1} \mathcal{K}^k 1 + \mathcal{K}^n w$. By induction, using the Beta integral,

$$(\mathcal{K}^k 1)(t) = (b\,\Gamma(\beta))^k \frac{t^{k\beta}}{\Gamma(k\beta + 1)},$$

whence $\sum_{k \geq 0} \mathcal{K}^k 1 = E_\beta(b\Gamma(\beta) t^\beta)$, which converges for every t. It remains to show $\mathcal{K}^n w \to 0$ uniformly on I: since w is continuous, $\|\mathcal{K}^n w\|_{L^\infty(I)} \leq \|w\|_{L^\infty(I)} (b\Gamma(\beta))^n T^{n\beta} / \Gamma(n\beta + 1) \to 0$ by the same computation. Passing to the limit gives the claim, and monotonicity of E_β on $[0, \infty)$ gives the uniform bound with $C(b, T, \alpha_+) = E_\beta(b\Gamma(\beta) T^\beta)$. □

Appendix B. Implementation Notes

Nodes. The LGR nodes on $[0, T]$ are obtained as the roots of $P_N + P_{N+1}$ on $[-1, 1)$ together with -1, mapped by $t = (x + 1)T/2$.

Assembly of $\mathbf{D}_N$. Rather than forming $\mathbf{Q}\mathbf{Y}\mathbf{Q}^{-1}$, we evaluate (11) directly, summing over k in increasing order and computing $\Gamma(k+1)/\Gamma(k+1-\alpha(t_j))$ through $\exp(\ln\Gamma(k+1) - \ln\Gamma(k+1-\alpha(t_j)))$ to avoid overflow. The shifted Legendre values are generated from the three-term recurrence, never from the monomial coefficients.

Quadrature. Every integral in the scheme involves either $t^{\alpha(0)}$ or $t^{k-\alpha(t)}$, both of which have an algebraic singularity in some derivative at $t = 0$. A plain Gauss rule converges only algebraically on such integrands: 400 points still leave approximately 10^{-6} in the projection coefficients of $t^{0.6}$, an error that would be indistinguishable from a discretisation error in a convergence table. We use a composite Gauss–Legendre rule on geometrically graded subintervals accumulating at the origin, which reaches machine precision with about 260 points.

Scaling. The rows of $\mathbf{D}_N$ differ in magnitude by several orders; we row-equilibrate the equality constraints of (P_N) before passing them to the solver, which materially improves SQP convergence.

Initialisation. $\mathbf{B} = 0$ with $\mathbf{A}$ the least-squares fit to the uncontrolled trajectory is more reliable than a zero start, which lies on the boundary of the feasible set when $u_{\min} = 0$. For nonconvex plants, continuation in the polynomial degree with re-projection between degrees is used in addition.

Verification. Every reported solution is checked by evaluating the continuous residual ${}^{C}D_{0^+}^{\alpha(t)} x_N - f(t, x_N, u_N)$ on a grid ten times finer than the collocation grid, and the constraints on the same grid, as required by step 7 of Algorithm 1. Following Proposition 7.4, the residual is reported on $[t_1, T]$ and, separately, on the full interval including the boundary layer.

Data and Code Availability. The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgements. The researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University (https://www.qu.edu.sa) for financial support (QU-APC-2026).

Authors' Contributions. All authors have read and approved the final version of the manuscript. The authors contributed equally to this work.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S.G. Samko, B. Ross, Integration and Differentiation to a Variable Fractional Order, Integral Transf. Spec. Funct. 1 (1993), 277–300. https://doi.org/10.1080/10652469308819027.

[2] C.F. Lorenzo, T.T. Hartley, Variable Order and Distributed Order Fractional Operators, Nonlinear Dyn. 29 (2002), 57–98. https://doi.org/10.1023/A:1016586905654.

[3] C.F.M. Coimbra, Mechanics with Variable-Order Differential Operators, Ann. Phys. 515 (2003), 692–703. https://doi.org/10.1002/andp.200310032.

[4] S. Patnaik, J.P. Hollkamp, F. Semperlotti, Applications of Variable-Order Fractional Operators: A Review, Proc. R. Soc. A 476 (2020), 20190498. https://doi.org/10.1098/rspa.2019.0498.

[5] H. Sun, A. Chang, Y. Zhang, W. Chen, A Review on Variable-Order Fractional Differential Equations: Mathematical Foundations, Physical Models, Numerical Methods and Applications, Fract. Calc. Appl. Anal. 22 (2019), 27–59. https://doi.org/10.1515/fca-2019-0003.

[6] O.P. Agrawal, A General Formulation and Solution Scheme for Fractional Optimal Control Problems, Nonlinear Dyn. 38 (2004), 323–337. https://doi.org/10.1007/s11071-004-3764-6.

[7] R. Kamocki, Pontryagin Maximum Principle for Fractional Ordinary Optimal Control Problems, Math. Methods Appl. Sci. 37 (2014), 1668–1686. https://doi.org/10.1002/mma.2928.

[8] M. Bergounioux, L. Bourdin, Pontryagin Maximum Principle for General Caputo Fractional Optimal Control Problems with Bolza Cost and Terminal Constraints, ESAIM Control Optim. Calc. Var. 26 (2020), 35. https://doi.org/10.1051/cocv/2019021.

[9] H. Wang, X. Zheng, Analysis and Numerical Solution of a Nonlinear Variable-Order Fractional Differential Equation, Adv. Comput. Math. 45 (2019), 2647–2675. https://doi.org/10.1007/s10444-019-09690-0.

[10] H. Ait Mohammed, M.E.-H. Mezabia, B. Tellab, A. Amara, H. Emadifar, Analysis of Caputo Fractional Variable Order Multi-Point Initial Value Problems: Existence, Uniqueness, and Stability, Bound. Value Probl. 2024 (2024), 127. https://doi.org/10.1186/s13661-024-01943-2.

[11] M.S. Souid, Z. Bouazza, S. Sabit, K. Sitthithakerngkiet, Numerical and Analytical Investigations of Hadamard Variable Order Fractional Differential Equations via Cumulative Distribution Functions, Fractal Fract. 10 (2026), 459. https://doi.org/10.3390/fractalfract10070459.

[12] M.H. Heydari, Z. Avazzadeh, A New Wavelet Method for Variable-Order Fractional Optimal Control Problems, Asian J. Control 20 (2018), 1804–1817. https://doi.org/10.1002/asjc.1687.

[13] F. Mohammadi, H. Hassani, Numerical Solution of Two-Dimensional Variable-Order Fractional Optimal Control Problem by Generalized Polynomial Basis, J. Optim. Theory Appl. 180 (2019), 536–555. https://doi.org/10.1007/s10957-018-1389-z.

[14] H. Hassani, Z. Avazzadeh, J.A. Tenreiro Machado, E. Naraghirad, A New Hybrid Method for Two Dimensional Nonlinear Variable Order Fractional Optimal Control Problems, Asian J. Control 23 (2021), 2004–2018. https://doi.org/10.1002/asjc.2351.

[15] Z. Avazzadeh, H. Hassani, A. Bayati Eshkaftaki, M.J. Ebadi, P. Agarwal, Optimal Solution of Nonlinear 2D Variable-Order Fractional Optimal Control Problems Using Generalized Bessel Polynomials, J. Vib. Control 31 (2025), 1457–1471. https://doi.org/10.1177/10775463241227475.

[16] H. Dehestani, Y. Ordokhani, M. Razzaghi, Piecewise Numerical Approach for Solving Piecewise Fractional Optimal Control and Variational Problems, Numer. Algorithms 101 (2026), 823–846. https://doi.org/10.1007/s11075-025-02024-w.

[17] P. Dai, X. Zheng, Y. Li, A Fast Iterative Method for Variable-Order Time-Fractional Optimal Control Problem, Numer. Linear Algebra Appl. 32 (2025), e70050. https://doi.org/10.1002/nla.70050.

[18] S. Soradi-Zeid, M. Alipour, A New Method for Solving Optimal Control Problems of Nonlinear Variable-Order Fractional Volterra–Fredholm Integro Differential Equations, Iran. J. Sci. Technol. Trans. Electr. Eng. 49 (2025), 1561–1577. https://doi.org/10.1007/s40998-025-00854-x.

[19] Z. Pirouzeh, K. Nassiri Pirbazari, M.H. Noori Skandari, Approximate Solution of Variable-Order Fractional Optimal Control Problems, Jordan J. Math. Stat. 19 (2026), 181–201. https://jjms.yu.edu.jo/index.php/jjms/article/view/1799.

[20] G. Elnagar, M.A. Kazemi, M. Razzaghi, The Pseudospectral Legendre Method for Discretizing Optimal Control Problems, IEEE Trans. Autom. Control 40 (1995), 1793–1796. https://doi.org/10.1109/9.467672.

[21] D. Garg, M. Patterson, W.W. Hager, A.V. Rao, D.A. Benson, et al., A Unified Framework for the Numerical Solution of Optimal Control Problems Using Pseudospectral Methods, Automatica 46 (2010), 1843–1851. https://doi.org/10.1016/j.automatica.2010.06.048.

[22] W. Kang, Rate of Convergence for the Legendre Pseudospectral Optimal Control of Feedback Linearizable Systems, J. Control Theory Appl. 8 (2010), 391–405. https://doi.org/10.1007/s11768-010-9104-0.

[23] W.W. Hager, J. Liu, S. Mohapatra, A.V. Rao, X.-S. Wang, Convergence Rate for a Gauss Collocation Method Applied to Constrained Optimal Control, SIAM J. Control Optim. 56 (2018), 1386–1411. https://doi.org/10.1137/16M1096761.

[24] W.W. Hager, H. Hou, S. Mohapatra, A.V. Rao, X.-S. Wang, Convergence Rate for a Radau *hp* Collocation Method Applied to Constrained Optimal Control, Comput. Optim. Appl. 74 (2019), 275–314. https://doi.org/10.1007/s10589-019-00100-1.

[25] T.A. Khalid, F. Alnoor, E. Babeker, E. Ahmed, A. Mustafa, Legendre Polynomials and Techniques for Collocation in the Computation of Variable-Order Fractional Advection–Dispersion Equations, Int. J. Anal. Appl. 22 (2024), 185. https://doi.org/10.28924/2291-8639-22-2024-185.

[26] T.A. Khalid, N.E. Taha, M.Y.A. Juma, M. Elmahi, N.H. Hagabdulla, et al., Order Modulation for Chaos Control and Hybrid Synchronization in a Variable-Order Fractional Arneodo System: Spectral Stability and Numerical Validation, Fractal Fract. 10 (2026), 376. https://doi.org/10.3390/fractalfract10060376.

[27] T. Binazadeh, E. Jafari, Synchronization of Chaotic Fractional-Order Systems with Input Saturation and Non-Vanishing Perturbations: Ultimate Boundedness Analysis and Transient Behavior Improvement, Int. J. Dyn. Control 13 (2025), 134. https://doi.org/10.1007/s40435-025-01633-6.

[28] S. Liu, R. Luo, L. Han, J. Li, Z. Meng, Fixed-Time Saturated Synchronization of a Class of Fractional-Order Chaotic Systems with Disturbances, Phys. Scr. 100 (2025), 105215. https://doi.org/10.1088/1402-4896/ae0ed2.

[29] S. Mirzajani, Control of Fractional-Order Chaotic Systems under Perturbations, Comput. Methods Differ. Equ. 13 (2025), 1037–1046. https://doi.org/10.22034/cmde.2025.59973.2556.

[30] T.A. Khalid, N.E. Taha, Quadratic Lyapunov Estimates for Variable-Order Fractional Derivatives: Kernel Monotonicity and Definition-Dependence, Mathematics 14 (2026), 3486. https://doi.org/10.3390/math14193486.

[31] T.A. Khalid, Application of Elzaki Transform Decomposition Method in Solving Time-Fractional Sawada–Kotera–Ito Equation, Malays. J. Math. Sci. 19 (2025), 691–706. https://doi.org/10.47836/mjms.19.2.17.

[32] R. Gorenflo, S. Vessella, Abel Integral Equations: Analysis and Applications, Lecture Notes in Mathematics 1461, Springer, Berlin, 1991. https://doi.org/10.1007/BFb0084665.

[33] Y. Li, Y. Chen, I. Podlubny, Mittag–Leffler Stability of Fractional Order Nonlinear Dynamic Systems, Automatica 45 (2009), 1965–1969. https://doi.org/10.1016/j.automatica.2009.04.003.

[34] S. Chen, J. Shen, L.-L. Wang, Generalized Jacobi Functions and Their Applications to Fractional Differential Equations, Math. Comput. 85 (2016), 1603–1638. https://doi.org/10.1090/mcom3035.

[35] M.A. Zaky, An Accurate Spectral Collocation Method for Nonlinear Systems of Fractional Differential Equations and Related Integral Equations with Nonsmooth Solutions, Appl. Numer. Math. 154 (2020), 205–222. https://doi.org/10.1016/j.apnum.2020.04.002.

[36] S. Erfani, S. Javadi, E. Babolian, An Efficient Collocation Method with Convergence Rates Based on Müntz Spaces for Solving Nonlinear Fractional Two-Point Boundary Value Problems, Comput. Appl. Math. 39 (2020), 260. https://doi.org/10.1007/s40314-020-01302-8.

[37] J.G. Lu, Chaotic Dynamics and Synchronization of Fractional-Order Arneodo's Systems, Chaos Solitons Fractals 26 (2005), 1125–1133. https://doi.org/10.1016/j.chaos.2005.02.023.

[38] T.A. Khalid, M.A.M. Ashmaig, H.M.E. Ahmed, B.A.A. Mahmoud, N.E. Taha, A Mesh-Independent Adjoint Consistency Defect in Optimal Control of the Caputo Time-Fractional Lindblad Equation: Sharp Classical-Limit Rate, Correction, and Convergence, Fractal Fract. 10 (2026), 619. https://doi.org/10.3390/fractalfract10090619.

[39] T.A. Khalid, N.E. Taha, M.Y. Juma, M.A.M. Ashmaig, M. Elmahi, et al., A Projection-Based Framework for Exact Partial Controllability of Semilinear Stochastic Fractional Evolution Equations with State-Dependent Delays, Impulses, and Q-Wiener Perturbations, Math. Comput. Appl. 31 (2026), 134. https://doi.org/10.3390/mca31040134.