Linear Regression with Truncated Path Signature

Niklas Weber

LMU Munich

July 26, 2023

Outline

The Path Signature

- ...is a sequence of iterated integrals.
- $\mathbb{D}$ is a time interval $[a, b] \in \mathbb{R}$
- Let $X : \mathbb{D} \to \mathbb{R}^d$ a $d \in \mathbb{N}$ dimensional continuous path
- Let X be of bounded variation
- Signature term of the multi-index $(i_1, ..., i_k)$ of length $k \in \mathbb{N}$, $(i_1, ...i_k) \subseteq \{1, ..., d\}^k$ is defined as the iterated (Riemann-Stieltjes) integral:

$$S(X)^{(i_1,\dots,i_k)} := \int\cdots\int_{a\leq t_1\leq \dots \leq t_k \leq b} dX^{i_1}_{t_1} \dots dX^{i_k}_{t_k}.$$

Subsequently the complete signature of such a path is defined as the sequence of all the signature terms of multi-indices with increasing length:

$$\begin{aligned} S(X) = (1, S(X)^1, ..., S(X)^d, S(X)^{(1,1)}, \\ S(X)^{(1,2)}, ..., S(X)^{(d,d)}, S(X)^{(1,1,1)}, ...) \end{aligned}$$

Truncated Signature

- Signature of X truncated at level m:

$$S^m(X) := (1, S(X)^1, ..., S(X)^{\overbrace{(d, ..., d)}^{\text{length } m}})$$

- Number of m-th order terms for d-dimensional path: d^m
  Therefore for $d \geq 2$, $S^m(X)$ would have

$$s_d(m) := \sum_{k=0}^{m} d^k = \frac{d^{m+1} - 1}{d - 1} \quad \text{terms.}$$

Truncated Signature II

	$d=2$	$d=3$	$d=6$
$m=1$	2	3	6
$m=2$	6	12	42
$m=5$	62	363	9330
$m=7$	254	3279	335922

Figure: Number of terms for typical values of m and d, [9].

- Decaying norm of terms (Lyons, 2014 [4], Fermanian, 2020 [9]): Let $X : [0,1] \to \mathbb{R}^d$ be a bounded variation path. Then for any $m \geq 0$,

$$\|S^m(X)\| \leq \sum_{k=0}^{m} \frac{\|X\|_{TV}^k}{k!} \leq e^{\|X\|_{TV}}.$$

Path Signature Example 1, Chevyrev [5]

$X : [a,b] \mapsto \mathbb{R}$, $X_t = t$. $(dX_t = \underbrace{\dot{X}_t}_{=1} dt)$

$$S(X)^0 = 1,$$

$$S(X)^1 = \int_a^b dX_t^1 = X_b^1 - X_a^1 = \frac{b-a}{1!},$$

$$S(X)^{(1,1)}_{a,b} = \iint_{a \le t_1 \le t_2 \le b} dX_{t_1}^1 \, dX_{t_2}^1 = \int_{a \le t_2 \le b} \underbrace{\int_{a \le t_1 \le t_2} dX_{t_1}^1}_{=S(X)^1_{a,t_2}} dX_{t_2}^1$$

$$= \int_{a \le t_2 \le b} S(X)^1_{a,t_2} \, dX_{t_2}^1 = \int_{a \le t_2 \le b} (t_2 - a) \, dX_{t_2}^1$$

$$= \int_{a \le t_2 \le b} (t_2 - a) \underbrace{\dot{X}_{t_2}}_{=1} dt_2 = \frac{(b-a)^2}{2!},$$

$$S(X)^{(1,1,1)}_{a,b} = \frac{(b-a)^3}{3!}$$

Path Signature Example 2, Chevyrev [5]

$X_t = (X_t^1, X_t^2) = (3+t, (3+t)^2) \quad t \in [0,5], \ (a=0, b=5)$
$dX_t = (dX_t^1, dX_t^2) = (dt, 2(3+t)dt).$

$$S(X)_{0,5}^{(1)} = \int_0^5 dX_t^1 = X_5^1 - X_0^1 = 8 - 3 = 5$$

$$S(X)_{0,5}^{(2)} = \int_0^5 dX_t^2 = X_5^2 - X_0^2 = 64 - 9 = 55$$

$$S(X)_{0,5}^{(1,1)} = \iint_{0 \le t_1 \le t_2 \le 5} dX_{t_1}^1 dX_{t_2}^1 = \int_0^5 \int_0^{t_2} dt_1 dt_2 = \int_0^5 t_2 dt_2 = \frac{25}{2}$$

$$S(X)_{0,5}^{(1,2)} = \iint_{0 \le t_1 \le t_2 \le 5} dX_{t_1}^1 dX_{t_2}^2 = \int_0^5 \int_0^{t_2} dt_1 \, 2(3+t_2) \, dt_2$$

$$= \int_0^5 6t_2 + 2t_2^2 \, dt_2 = \frac{475}{3}$$

Properties I

- $\tilde{X} = X + a \implies S(\tilde{X}) = S(X)$
- Invariance under time-reparametrization

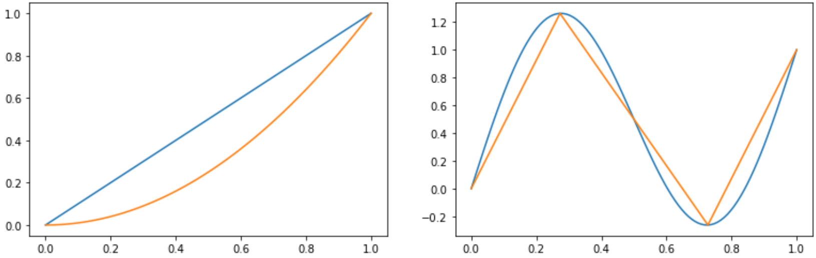

Figure: Time-reparametrizations have the same signature

- $\implies$ Add monotone time dimension. Augmented paths are better distinguishable

Properties II

Which information about the path is captured by the signature?

- 1st order terms: Increments
- 2nd order terms: Areas outlined by path

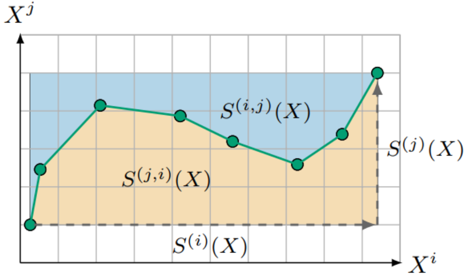

Figure: Interpretation of 2nd order signature terms, Fermanian [9].

- $\implies S(X)^{(i)}S(X)^{(j)} = S(X)^{(i,j)} + S(X)^{(j,i)}$
- Higher order terms: Info about joint evolution of tuples of coordinates

Properties III

- Chevyrev, Kormilitzin [5]: "A natural question one may ask is the following: is a path completely determined by its signature? [...] the answer, in general, is no."
- "For example, one can never recover from the signature the exact speed at which the path is traversed, nor can one tell apart the signature of a trivial constant path and that of a path concatenated with its time-reversal."
- "However, [...] this is essentially the only information one loses from the signature. For example, for a path X which never crosses itself, the signature is able to completely describe the image and direction of traversal of the path (that is, all the points that X visits and the order in which it visits them)."

Theorem (Hambly, Lyons [3], Fermanian [9])

Assuming $X \in BV(\mathbb{R}^d)$ contains one monotone coordinate, then $S(X)$ characterizes X up to translations and reparametrizations.

Computational considerations I, [8]

- Linear path: $X_t = (X_t^1, X_t^2) = (a_1 + b_1 t, a_2 + b_2 t) \quad [s,t] \subset \mathbb{R}$

$$
\begin{aligned}
S(X)_{s,t}^{(i)} &= \int_s^t dX_u^i = b_i(t-s) \\
S(X)_{s,t}^{(1,1)} &= \iint_{s \leq u_1 \leq u_2 \leq t} dX_{u_1}^1 \, dX_{u_2}^1 = \int_s^t \int_s^{u_2} b_1^2 \, du_1 \, du_2 \\
&= b_1^2 \int_s^t (u_2 - s) \, du_2 = \frac{b_1^2 (t-s)^2}{2}
\end{aligned}
$$

$$
S(X)_{s,t}^{(i_1,\ldots,i_k)} = \frac{b_{i_1} \ldots b_{i_k} (t-s)^k}{k!}
$$

Computational considerations II

Theorem (Chen's Theorem [1], Lyons et al. [2], Fermanian [8])

Let $X : [s, t] \to \mathbb{R}^d$ and $Y : [t, u] \to \mathbb{R}^d$ be two paths with bounded variation. Then for any multi-index $(i_1, ..., i_k) \subset \{1, ..., d\}^k$,

$$S(X * Y)^{(i_1,\dots,i_k)} = \sum_{l=0}^{k} S(X)^{(i_1,\dots,i_l)} \cdot S(Y)^{(i_{l+1},\dots,i_k)}$$

- Paths on the computer are always points.
- Interpolate linearly
- Calculate signature terms piecewise. (No integration necessary)
- Concatenate with Chen's Formula.
- $\rightarrow$ Python *iisignature*, Reizenstein and Graham [6].

Motivation: Linear Regression on Path Signature

The next theorem by Király, Oberhauser (2019)...

Theorem (Király, Oberhauser [7], Fermanian [9])

Let f be a continuous function $f \in C(K, \mathbb{R})$ on a compact set of bounded variation paths $K \subset BV(\mathbb{R}^d)$ with at least one monotone coordinate and $X_0 = 0$. For any $\epsilon > 0$ there exists $m^* \in \mathbb{N}$ and $\beta^* \in \mathbb{R}^{s_d(m^*)}$ such that

$$\sup_{X \in K} \| f(X) - \langle \beta^*, S^{m^*}(X) \rangle \| < \epsilon.$$

...motivates the following idea. We want to predict the real random variable $Y \in \mathbb{R}$ using functional covariate $X \in \mathbb{R}^d$.

Signature linear model, Fermanian [9]

$$\mathbb{E}\left[Y | X \right] = f(X) = \langle \beta^*_{m^*}, S^{m^*}(X) \rangle.$$

Truncation level $m^* \Rightarrow$ We need to choose a truncation level.

Estimating the truncation order

- Data: $D_n = \{(X_1, Y_1), ..., (X_n, Y_n)\}$
- Approach: Penalized empirical risk minimization.
- $\hat{m}$ is chosen by

$$\hat{m} = \min\left(\underset{m\in\mathbb{N}}{\operatorname{argmin}}(\hat{L}_n(m) + pen_n(m))\right).$$

- pen is a penalization function.
- Empirical risk and empirical minimal risk:

$$\hat{R}_{m,n}(\beta) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle\beta, S^m(X_i)\rangle)^2$$

$$\hat{L}_n(m) = \min_{\beta\in B_{m,\alpha}} \hat{R}_{m,n}(\beta) = \hat{R}_{m,n}(\hat{\beta}).$$

Performance bound

Under some assumptions it holds that

Theorem, Fermanian [9]

For any $n \geq n_0$

$$\mathbb{P}(\hat{m} \neq m^*) \leq C_1 \exp(-C_2 n^{1-2\rho}),$$

where $0 < \rho < \frac{1}{2}$, and the constants C_1 and C_2 are defined by...

Assumptions

- $\text{pen}_n(m) = K_{\text{pen}} n^{-\rho} \sqrt{s_d(m)}$
- $\beta^*_{m^*}$ lies inside α-Ball
- there exist two real numbers $K_Y > 0$ and $K_X > 0$ such that almost surely $|Y| \leq K_Y$ and $||X||_{TV} \leq K_X$.
- ...

Proof Idea I

- Goal: bound probability of choosing wrong truncation order:

$$\mathbb{P}(\hat{m} \neq m^*) \leq \dots$$

- We split the probability in two sums.

$$\mathbb{P}(\hat{m} \neq m^*) = \sum_{m > m^*} \mathbb{P}(\hat{m} = m) + \sum_{m < m^*} \mathbb{P}(\hat{m} = m).$$

- We find upper bounds of the form

$$\mathbb{P}(\hat{m} = m) \leq ...$$

- They will be derived by establishing a relation between

$$\mathbb{P}(\hat{m} = m) \quad \text{and} \quad \mathbb{P}(\sup_{\beta \in B_{m,\alpha}} |\hat{R}_{m,n}(\beta) - R_m(\beta)| \geq \dots)$$

Proof Idea II

- Showing that

$$Z_{m,n}(\beta) := \hat{R}_{m,n}(\beta) - R_m(\beta)$$

is a separable, subgaussian process...

- ...enables us to use an inequality of the form

$$\mathbb{P}\left(\sup_{\beta \in B_{m,\alpha}} (Z_{m,n}(\beta) - Z_{m,n}(\beta_0)) \geq \dots \right) \leq \dots \quad \text{for any } \beta_0.$$

See Van Handel, Probability in high dimension [11].

Question?

- We know Signature Regression is an interesting alternative to functional regression (Fermanian [9])
  $\rightarrow$ less assumptions, relatively good in high dimensions (despite computational cost)
- Is it an alternative for Linear Regression (on the path itself)
- Does this depend on how fine/long the path is?
- Is Signature Regression suitable for Credit Cycle Forecasting?

Algorithm

Algorithm 1 Signature Regression

Get or generate **Data** $\{(X_1, Y_1), ..., (X_n, Y_n)\}$
Add time dimension (and interpolate **Data** if necessary)
Split **Data** into **Train** and **Test** set
procedure SELECT $\hat{m}$ VIA CV(**Train**)
    Calculate m_{max} ▷ Such that $s_d(m) \leq 10.000$
    for $0 \leq m \leq m_{max}$ **do**
        Split **Train** in five subsets **Train**$_j$ ▷ for cross-validation
        Fit Ridge Regression to $\{S^m(X_i), Y_i\}$ in **Train**$_j$ ▷ Ridge-α by CV
        Measure average performance on **Train**$_j$
    end for
    Choose best performing m as $\hat{m}$
end procedure
procedure COMPARE REGRESSION TYPES(**Train**, **Test**, $\hat{m}$)
    Fit Ridge Regression to $\{S^{\hat{m}}(X_i), Y_i\}$ in **Train** ▷ Here we get get $\hat{\beta}^{\hat{m}}$
    With $\hat{m}$ and $\hat{\beta}^{\hat{m}}$ predict $\{\hat{Y}_i\}$ from $\{S^{\hat{m}}(X_i)\}$ in **Test**
    Measure prediction performance of Signature Regression (e.g. MSE, R,...)
    Reshape X_i into one long vector $\tilde{X}_i$
    Fit Ridge Regression to $\{\tilde{X}_i, Y_i\}$ in **Train**
    Predict $\{\hat{Y}_i\}$ from $\{\tilde{X}_i\}$ in **Test**
    Measure prediction performance of Linear Regression (e.g. MSE, R,...)
end procedure

Figure: Algorithm

Experiment I

- Path: For $1 \leq i \leq n$, let $X_i : [0,1] \to \mathbb{R}^d$, $X_{i,t} = (X_{i,t}^1, ..., X_{i,t}^d)$ be defined by

$$X_{i,t}^k = \alpha_{i,1}^k + 10\alpha_{i,2}^k \sin\left(\frac{2\pi t}{\alpha_{i,3}^k}\right) + 10(t - \alpha_{i,4}^k)^3, \quad 1 \leq k \leq d, \tag{1}$$

  where $\alpha_{i,l}^k$, $1 \leq l \leq 4$ are sampled uniformly on $[0,1]$.
- Response: For some m^*

$$Y_i = \langle \beta, S^{m^*}(X_i) \rangle + \epsilon_i, \tag{2}$$

  where ϵ_i uniformly on [-100,100], and β is given by

$$\beta_j = \frac{1}{1000} u_j, \quad 1 \leq j \leq s_d(m^*), \tag{3}$$

  with u_j sampled uniformly on [0,1].

Experiment I

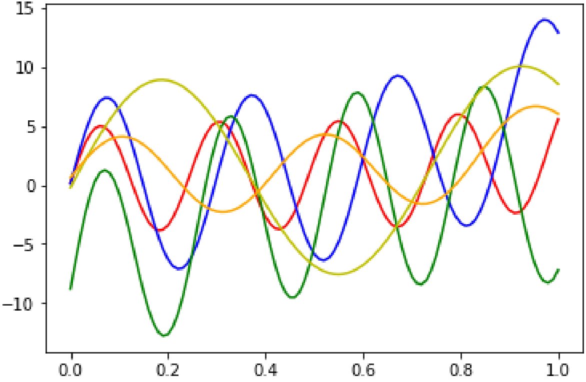

Figure: Example paths

Experiment I

- Choose m^*=5, $d = 3$
- Consider $num \in [3, 5, 10, 20, 50, 100]$ and $nPaths \in [33, 50, 100, 200, 500, 1000]$
- Run every configuration 20 times
- We analyse $\hat{m}$ and $R^2 = 1 - \frac{u^2}{\sigma^2}$
- $R^2 = 1$: perfect prediction
- $R^2 > 0$: our mean square error is lower than the variance of the responses. We are better than always guessing the average response.
- $R^2 < 0$: our predictions are worse than guessing the average. Our model might not add valuable information.

Experiment I

- $\hat{m}$

1.2	1.45	2.5	2.15	2.7	2.3
2.25	2.25	2.1	2.25	2.25	2.3
2.35	2.6	2.95	2.4	2.75	2.6
3.05	2.9	3.35	2.5	3.65	3.35
3.25	3.25	3.35	4.1	3.95	4.8
3.15	3.1	4.2	4.95	5.05	4.85

Figure: $\hat{m}$ average of signature regression

1.36382	1.98683	1.93649	1.98179	2.30434	2.07605
1.57718	1.17792	1.8412	2.27761	2.21077	2.05183
1.10793	1.49666	1.98683	2.00998	1.78536	1.90788
0.864581	0.888819	1.4239	1.93649	1.82414	1.85135
0.433013	0.433013	1.4586	1.37477	1.68745	1.249
0.357071	0.43589	1.07703	0.384057	0.384057	0.726292

Figure: $\hat{m}$ std of signature regression

Experiment I

- R^2

-0.621198	-4.97949	-0.626383	-0.883898	-1.04474	-1.17464
-0.0747186	-0.841111	-0.75699	-0.802904	-1.48537	-2.1868
0.0207454	-0.226368	-0.581713	-0.690532	-1.58578	-3.19478
0.0872331	0.0350942	-0.118908	-0.886698	-1.3603	-1.08825
0.210323	0.168361	0.0376061	-0.23387	-0.704523	-0.732267
0.232846	0.185209	0.0879272	-0.00666589	-0.285289	-1.74515

Figure: R^2 average of linear regression

-0.160823	-0.145855	-0.0414059	-0.400529	-0.28341	-0.182562
-0.183304	-0.271865	-0.0371353	-0.297478	-0.216546	-0.171845
0.0447583	-0.104427	-0.056792	-0.110745	-0.21987	-0.475177
0.159418	0.0903768	0.0899578	0.0366832	-0.049195	0.0844227
0.334665	0.271292	0.119611	0.107769	0.145808	0.185221
0.418354	0.333647	0.253306	0.411091	0.479729	0.332093

Figure: R^2 average of signature regression

Experiment II

- Path: Same as before $X_{i,t}$, $t \in [0,1]$.
- Response: $Y_i = \max(X^1_{i,T+\Delta t}, \ldots, X^d_{i,T+\Delta t})$.
- R^2

-0.275266	-2.61321	-0.333922	-0.395462	-0.408113	-0.125808
-0.210793	-0.41075	-0.596805	-0.304509	-0.227324	-0.411225
-0.0646064	-0.472868	-0.400313	-0.264366	-0.205634	-0.409057
0.0828974	-0.0169786	-0.0619481	-0.208777	0.0608	-0.0293006
0.0912928	0.0803163	0.0735291	0.0880184	0.216047	0.234262
0.128028	0.113594	0.156082	0.233046	0.195483	0.358812

Figure: R^2 average of linear regression

-4.40398	-4.53858	-3.87305	-2.2928	-3.1686	-3.72546
-2.67402	-2.34836	-2.88254	-0.645596	-2.32356	-2.13294
-2.61359	-1.66615	-1.65943	-0.0409393	-0.955918	-1.15959
-4.53347	-0.632873	-0.659369	0.151579	0.236565	0.245967
-0.178267	-0.0982401	0.00358971	0.26997	0.524609	0.657058
0.0138293	0.00640355	0.0985228	0.328635	0.576883	0.715937

Figure: R^2 average of signature regression

Experiment III

- Path: For $1 \leq i \leq n$, let $X_i : [0,1] \to \mathbb{R}^d$, $X_{i,t} = (X^1_{i,t}, ..., X^d_{i,t})$ be defined by

$$X^k_{i,t} = \alpha^k_i t + \epsilon^k_{i,t} \tag{4}$$

where α^k_i, is sampled uniformly on $[-3,3]$ and ϵ^k_i is a Gaussian process with exponential covariance matrix.

- Response:

$$Y_i = \|a_i\| \tag{5}$$

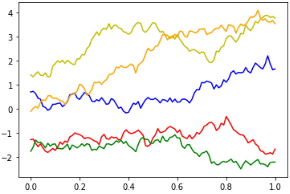

Figure: Example paths

Experiment III

- R^2

-0.683899	-0.548641	-0.479082	-0.88424	-1.32343	-1.13049
-0.181698	-0.240872	-0.305645	-0.338029	-0.385171	-0.519244
-0.136998	-0.118233	-0.128771	-0.124004	-0.26843	-0.249006
-0.0683324	-0.0447143	-0.0868118	-0.0930514	-0.140258	-0.112033
-0.0168486	-0.0343449	-0.0443118	-0.0438327	-0.0781499	-0.0826172
-0.0198128	-0.0212953	-0.0173746	-0.0287907	-0.0380603	-0.0553162

Figure: R^2 average of linear regression

-14.6591	-8.36975	-8.27958	-7.64022	-9.33557	-5.75476
-6.94528	-5.44725	-4.5728	-5.76347	-0.985052	-1.75431
-3.6532	-2.71803	-2.43551	-3.06354	-0.0907635	-0.148253
-0.374808	-0.416515	-0.550551	-0.746418	0.1278	0.155087
0.0527765	0.040163	0.0172433	0.0288435	0.234067	0.207195
0.16739	0.165341	0.181229	0.173795	0.254372	0.248059

Figure: R^2 average of signature regression

Credit Cycle Forecasting

- Financial institutions want to forecast how the credit cycle is going to develop. (Favourable or adverse environment)
- A proxy of such an indicator can be the probability of default for a region/sector.
- Response: Probability of default for North american firms from 1990 until 2021 published by the Credit Research Initiative of the National University of Singapore [10].
- For forecasting use US GDP growth, US unemployment, S&P 500 growth and US interest-rate-spread (i.e. long-term-interest-rate minus short-term-interest-rate).
- Try to predict next year PD with 3-year path of the predictors.
- Results not promising
- ToDo: Try again with new data (quarterly)

References I

[1] Kuo-Tsai Chen. "Integration of paths — a faithfuhl representation of paths by noncommutative formal power series". In: *Transactions of the American Mathematical Society* 89 (1958), pp. 395–407.

[2] Terry J Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer, 2007.

[3] Ben Hambly and Terry Lyons. "Uniqueness for the signature of a path of bounded variation and the reduced path group". In: *Annals of Mathematics* 171.1 (Mar. 2010), pp. 109–167. ISSN: 0003-486X. DOI: 10.4007/annals.2010.171.109. URL: http://dx.doi.org/10.4007/annals.2010.171.109.

[4] Terry Lyons. "Rough paths, signatures and the modelling of functions on streams". In: *arXiv preprint arXiv:1405.4537* (2014).

[5] Ilya Chevyrev and Andrey Kormilitzin. "A primer on the signature method in machine learning". In: *arXiv preprint arXiv:1603.03788* (2016).

References II

[6] Jeremy Reizenstein and Benjamin Graham. *The iisignature library: efficient calculation of iterated-integral signatures and log signatures*. 2018. arXiv: 1802.08252 [cs.DS].

[7] Franz J Király and Harald Oberhauser. "Kernels for sequentially ordered data". In: *Journal of Machine Learning Research* 20.31 (2019), pp. 1–45.

[8] Adeline Fermanian. *Embedding and learning with signatures*. 2020. arXiv: 1911.13211 [stat.ML].

[9] Adeline Fermanian. "Linear functional regression with truncated signatures". In: *arXiv preprint arXiv:2006.08442* (2020).

[10] The Credit Research Initiative - NUS Asian Institute of Digital Finance. *Aggregate PD & AS*. 2022. URL: https://nuscri.org/en/data/cdsaggregatedata/e501s0/0/ (visited on 02/02/2022).

References III

[11] Ramon van Handel. "Probability in high dimension, December 2016". In: APC.