

点過程を出力する線形因果モデル

統計数理研 尾形良彦

1. はじめに

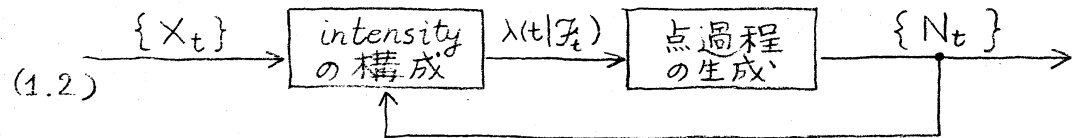
$\{\mathcal{F}_t, t \geq 0\}$ は t につれて増加する σ -加法族とする。これに対して adapted な非負の確率過程 $\{\lambda(t|\mathcal{F}_t), t \geq 0\}$ が与えられると、これを *intensity* とする点過程 (N_t, P) が唯一存在する。すなわち

$$(1.1) \quad P\{N_{t+\Delta t} - N_t = 1 \mid \mathcal{F}_t\} = \lambda(t|\mathcal{F}_t)\Delta t + o(\Delta t).$$

(このように *intensity* から点過程を生成する効率の良いアルゴリズムは「削ぎ落とし法」を応用すると得られる。[6]) ここでもし $\mathcal{F}_t = \sigma\{N_s, 0 \leq s < t\}$ のとき、この点過程は *self-exiting* であると呼ばれる。(*self-exiting* な点過程をデータとするときの尤度による推定方式の漸近理論は [5] に与えられている。)

本報告では、*intensity* 過程がもう一つの確率過程 $\{X_t, t \geq 0\}$ によって調節されている事態について考える。

このシステムを図示すると次のようになる。



input $\{X_t\}$ は点過程であっても良いし、または非負の連続過程 $\{x(s)\}$ の累積過程 $X_t = \int_0^t x(s)ds$ でも良い。 $\mathcal{F}_t = \sigma\{X_s, N_s; 0 \leq s < t\}$ すなわち $\{X_t\}$ と $\{N_t\}$ が観測データとなる。

このモデルの目的は2本のデータ $\{X_t\}$ と $\{N_t\}$ が与えられたときに、これらに統計的因果関係があるか否か、もしあるとすれば、どのようなものか、を見ることである。例えば、地震の系列 $\{N_t\}$ に対してあるデータ $\{X_t\}$ は前兆としての意味を持つのか、持つとすれば、どのくらいの遅れがあるか。2本の神経インパルス系列の関連性、等々、考えられる。

2. 線形因果モデル

点過程 (N_t, P) は充分小さな区間 $(t, t+\Delta t)$ に高々1個の点しか起らない (orderliness) と考えてよい、すなわち

$$(2.1) \quad P\{N_{t+\Delta t} - N_t \geq 2 \mid \mathcal{F}_t\} = o(\Delta t)。$$

すると (1.1) と (2.1) によって

$$(2.2) \quad E\{N_{t+\Delta t} - N_t | \mathcal{F}_t\} = \lambda(t|\mathcal{F}_t)\Delta t + o(\Delta t)$$

となり、これは $\mathcal{F}_t$ の情報のもとでの区間 $(t, t+\Delta t)$ が起るか否かについての最良予測になっている。 $\mathcal{F}_t$ は $\{X_t\}$, $\{N_t\}$ の過去の履歴となっているのだから、我々は最も簡単なモデル、すなわち最良予測が $\{X_t\}$, $\{N_t\}$ について線形に与えられる場合を考える。すなわち

$$(2.3) \quad \lambda(t|\mathcal{F}_t) = \mu + \int_0^t g(t-s) dN_s + \int_0^t h(t-s) dX_s,$$

ただし最後の項は、 $\{X_t\}$ が点過程か $x(s)$ の累積過程かによって、 $\sum_{x_i < t} h(t-x_i)$ か $\int_0^t h(t-s) x(s) ds$ となる。

システム (1.2) のもとで $\{X_t\}$ と $\{N_t\}$ に統計的因果関係がないことは $h(s) \equiv 0$ となることに対応する。 $g(s)$ は $\{N_t\}$ の *intensity* に対する *feedback* のあり方を示す。我々の興味は、したがって、応答関数たち $g(s)$, $h(s)$ の形状を知ることである。

intensity モデル (2.3) は形式的には、Hawkes [4] の *mutually-exciting* 点過程の *intensity* ベクトルの一成分と見なせないこともない ($\{X_t\}$ が点過程のとき)。しかしここでのモデルに於いて我々は *input* データ $\{X_t\}$ の確率法則 (統計的構造) は一切向わない。従って *input-output* の $\{X_t\}$ と $\{N_t\}$ の同時分布 (尤度) は全く不明である。

それにも拘らず、Cox [3] の意味での *partial* な対数尤度が

$$(2.4) \quad \mathcal{L}_T = \int_0^T \log \lambda_\theta(t | \mathcal{F}_t) dN_t - \int_0^T \lambda_\theta(t | \mathcal{F}_t) dt,$$

(但し $\lambda_\theta(t | \mathcal{F}_t)$ は (2.3) に与えたもの) と書くことができる。この *partial-likelihood* に基づく推定方式の漸近理論は、 $(\{X_t\}, \{N_t\})$ の定常エルゴード性等の正則条件のもとでは [5] に於けると類似に、保証される。従って *large sample* のもとでは、対数尤度の χ^2 -性、AIC の使用などが正当化される。

3. Parametrization と 尤度計算

尾崎および赤池 [7] は Hawkes の *self-exting* 過程の応答関数 $g(t)$ をラゲール型多項式

$$(3.1) \quad g(t) = \sum_{k=0}^K a_k t^k e^{-ct}$$

によって近似することを提案している。この型の応答関数は尤度関数における非線形最適化や *simulation* に際して *recursive* な構造をもっていて効率のよい計算を可能にする (赤池 [1], および [6])。これらから、もうひとつの応答関数についても

$$(3.2) \quad h(t) = \sum_{m=0}^M b_m t^m e^{-ct}$$

とパラメータ表現する。こうして時間区間 $[0, T]$ に input に点列 $\{x_j\}_{j=1, \dots, J}$, output に点列 $\{t_i\}_{i=1, \dots, I}$ を観測したとき partial log-likelihood (2.4) は次のとおり。

$$\begin{aligned}
 (3.3) \quad \mathcal{L}_T(\mu, c, a_0, \dots, a_K, b_0, \dots, b_M) \\
 = \sum_{i=1}^I \log \left\{ \mu + \sum_{k=0}^K a_k Q_k(i) + \sum_{k=0}^M b_k R_k(i) \right\} - \mu T \\
 - \sum_{i=1}^I \sum_{k=0}^K a_k S_k(T - t_i) - \sum_{j=1}^J \sum_{k=0}^M b_k S_k(T - x_j).
 \end{aligned}$$

ただし $Q_k(i)$, $R_k(i)$ そして $S_k(t)$ は以下のごとく recursive に書かれたものである。すなわち初期値として $Q_0(1) = R_0(1) = 0$ かつ $S_0(t) = (1 - e^{-ct})/c$ とおく。そして $k=0, 1, 2, \dots$, $i=0, 1, 2, \dots$ に対して順次

$$\begin{aligned}
 Q_k(i+1) &= (\Delta t_i)^k e^{-c \Delta t_i} + \sum_{l=0}^k \binom{k}{l} (\Delta t_i)^l e^{-c \Delta t_i} Q_l(i) \\
 (3.4) \quad R_k(i+1) &= D_k(t_i, t_{i+1}) + \sum_{l=0}^k \binom{k}{l} (\Delta t_i)^l e^{-c \Delta t_i} R_l(i) \\
 S_{k+1}(t) &= \{(k+1)S_k(t) - t^{k+1}e^{-ct}\}/c
 \end{aligned}$$

によって定義していく。ここに $t_0 = 0$, $\Delta t_i = t_{i+1} - t_i$,

$$D_k(t_i, t_{i+1}) = \sum_{t_i \leq x_j < t_{i+1}} (t_{i+1} - x_j)^k e^{-c(t_{i+1} - x_j)}$$

は、その都度計算される。

もし input が $X_t = \int_0^t x(s) ds$, $0 \leq t \leq T$ なる $\{x(s)\}$

が観測されるとする。このとき $\{x(s)\}$ は離散化されて X_t は次のように近似する。 $[0, T]$ を J 等分して $\sigma_j = j \cdot T/J$,

$$(3.5) \quad X_L = \sum_{j=1}^L x(\sigma_j) \cdot T/J$$

とする。こうして、この場合の partial-log-likelihood は

$$(3.6) \quad \begin{aligned} & \mathcal{L}_T(\mu, c, a_0, \dots, a_K, b_0, \dots, b_M) \\ &= \sum_{i=1}^I \log \left\{ \mu + \sum_{k=0}^K a_k Q_k(i) + \sum_{k=0}^M b_k R'_k(i) \right\} - \mu T \\ & \quad - \sum_{i=1}^I \sum_{k=0}^K a_k S_k(T-t_i) - \sum_{j=1}^J \sum_{k=0}^M b_k S'_k(j) \end{aligned}$$

で与えられる。ここで $Q_k(i)$ と $S_k(t)$ は (3.4) と同じに定義されるが、

$$(3.7) \quad \begin{aligned} R'_k(i+1) &= D'_k(t_i, t_{i+1}) + \sum_{l=0}^k \binom{k}{l} (\Delta t_i)^l e^{-c \Delta t_i} R'_l(i) \\ S'_k(j) &= S_k(T - \sigma_j) \cdot x(\sigma_j) \\ D'_k(t_i, t_{i+1}) &= \sum_{t_i \leq \sigma_j < t_{i+1}} x(\sigma_j) (t_{i+1} - \sigma_j)^k e^{-c(t_{i+1} - \sigma_j)} \end{aligned}$$

である。

関数たち (3.3), (3.6) の gradient ベクトル, Hessian 行列も recursive な量たち (3.4), (3.7) の有理結合で与えられるが、ここでは省略する。

応答関数を近似するラゲール型多項式の最適次数は AIC によって選択する。

4. 或る条件付情報量について

output データ $\{N_t\}$ に対して input データ $\{X_t\}$ の寄与の割合を示す条件付情報量は次のように定義される。

$$(4.1) \quad I_T(X|N) = E \left[\int_0^T \lambda(t|\mathcal{F}_t) \log \frac{\lambda(t|\mathcal{F}_t)}{\lambda(t|\mathcal{H}_t)} dt \right],$$

ただし $\mathcal{F}_t = \sigma\{N_s, X_s; 0 \leq s < t\}$ として $\mathcal{H}_t = \sigma\{N_s, 0 \leq s < t\}$ 。
いま X が点過程のとき、 X と N の相互情報量 (シャノンの意味での) $I_T(X; N)$ と条件付情報量の関係は

$$(4.2) \quad I_T(X; N) = I_T(X|N) + I_T(N|X)$$

で与えられることが知られている (Bremaud [2])。

もし $\{N_t\}$ の各 jump $\{t_i\}$ で $\lambda(t_i|\mathcal{F}_{t_i})$ と $\lambda(t_i|\mathcal{H}_{t_i})$ の値が、わかっているならば $I_1(X|N)$ の不偏一致推定量として

$$(4.3) \quad \hat{I}_1(X; N) = \frac{1}{T} \int_0^T \log \left\{ \lambda(t|\mathcal{F}_t) / \lambda(t|\mathcal{H}_t) \right\} dN_t$$

を考えることができる。しかしながら実際には、そのようなことは望めない。そこで線形因果モデル (2.3) を仮定してモデルを適合させて推定された $\hat{\lambda}(t|\mathcal{F}_t)$ および $\hat{\lambda}(t|\mathcal{H}_t)$ によって (4.3) に代入することが提案される。ただし $\hat{\lambda}(t|\mathcal{H}_t)$ は (2.3) において $h(s) \equiv 0$ を仮定したモデル (self-exciting) の推定によって得られる。

5. シミュレーションと群発地震データの解析

まず真の構造としてラゲール型多項式の応答関数を与えてそれから作ったデータをもとに3節で述べた尤度計算をして推定すると、もとの応答関数が復元できるだろうか。図1は真の応答関数のグラフを与えている。上段の2つは(1.2)において外からの入力に無関係な *self-exciting* の場合、下段の2つは外からの入力にも関係する場合である。左側の2つのグラフが *self-exciting* な部分、右側の2つが外からの入力に対する応答を示すものである。LGの次にある数字はグラフのラゲール型多項式の次数を示している。これら2つの場合のデータを「削ぎ落とし法」(〔5〕)で時間区間2000.0で生成した。入出力の2本のデータは生成されるときは *mutually-exciting* なものとして生成してもかまわないからそのようにした。図2は、図1のもとのデータについて最適な推定モデルのグラフである。それぞれのモデルのラゲール型多項式の9次までのAICの最小になる次数の組み合わせを選び、その係数は最尤推定値である。右側にそれぞれの4節で定義される情報量の推定値が与えてある。相互情報量 I は相関 ρ に対して $I = -\frac{1}{2} \log(1-\rho)$ (但し2次元正規分布の場合) の関係があるから大体の感じと矛盾しないであろう。このグラフから応答関数の復元について、2節の終りで述べていることを合

わせて確認された。

表1は長野県松代の群発地震の最頻期のデータである。
1965年10月1日頃に始まり4~5年続いたこの地震群のデータでマグニチュードが4.0前後以上のもので、定常に起っていると思われる90日~800日の間のデータをとった。従って1965年12月30日0時0分0秒を0.0としたときに続々と起った時点を記してある。上段のLMデータはマグニチュード4.6以上、下段は4.5以下のもので、時間1.0は1日をあらわ小数は24時間を10進法に展開した。

図3は(1.2)に於いてSMデータを入力、LMを出力としたとき、図4は逆にLMを入力、SMを出力とした、応答関数のAIC最小のもののグラフである。それぞれの上段がself-excitingな部分、下段が入力による応答を示している。図3の上段の応答で負になっているのは、大きい地震が起ると次にまた大きい地震が起りにくくなっていることを示している。また下段のグラフたちを見比べると小さい地震が大きい地震に与える応答よりも逆の応答の方が強いことを示している。横座標の数字は1.0が1日を意味するので応答の時間経過がだいたいどのようになっているのかを示している。

6. むすびと謝辞

この報告で提案された線形因果モデルは2本のデータ（うち1つは点過程）の時間的因果関係のあり方を推定された応答関数のグラフをとおして見、予測するのに役だてようとする一つの現実的な方法と考えられる。入力データは複数にできるし、(2.3)における応答関数の他の *Parameterization* も考えられるが、それはまた別の機会に発表する。このモデルの尤度計算やシミュレーションのためのプログラムパッケージも近い将来に発表する予定である。

表1に与えられた地震データは、1976年に学術振興会によってニュージーランド、ヴィクトリア大学の Vere-Jones 教授を招いた際に於ける著者を含む統計数理研究所の共同プロジェクト（代表者、赤池弘次氏）に、気象庁より提供されたものに基づいている。この点について深く感謝する。また、ここで提案されたモデルの基本的なアイディアは、赤池弘次氏や尾崎統氏に示唆されたものである。尤度関数の非線形最適化のプログラムは TIMSAC 78 のものを使った。この際、北川源四郎氏には、様々の貴重な援助をいただいた。さらに尤度計算のプログラム作成にあたっては桂康一氏に一貫した援助をしていただいた。以上の統計数理研究所の各氏に深く感謝する。京大防災研の尾池助教授には、地震データなどに関してたいへん

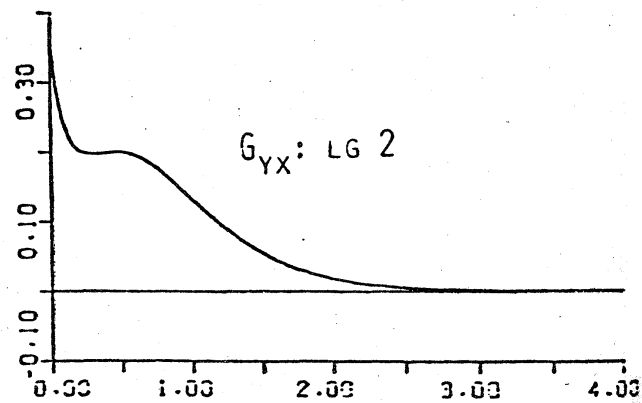
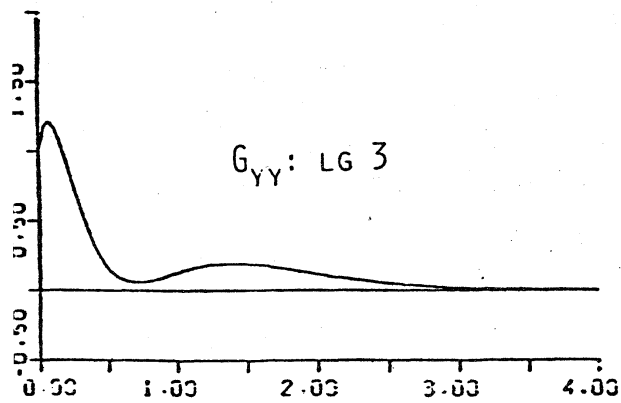
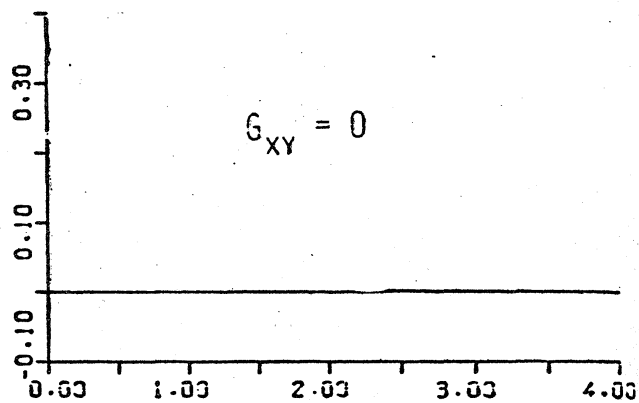
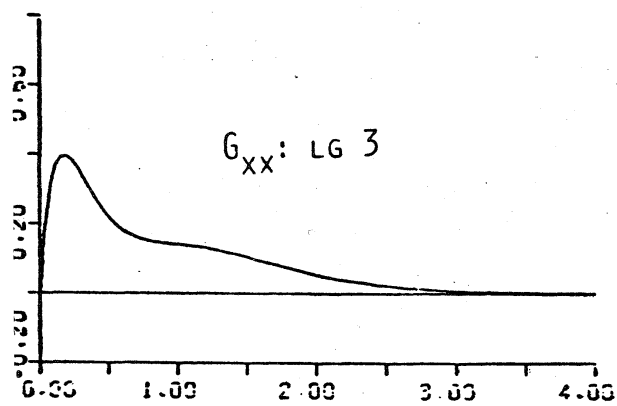
有益な助言をいただいた。

REFERENCES

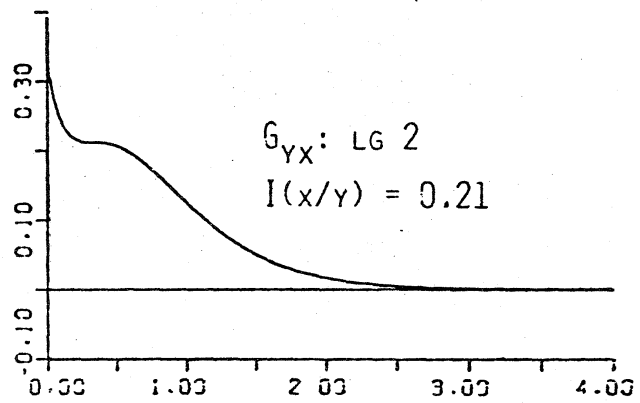
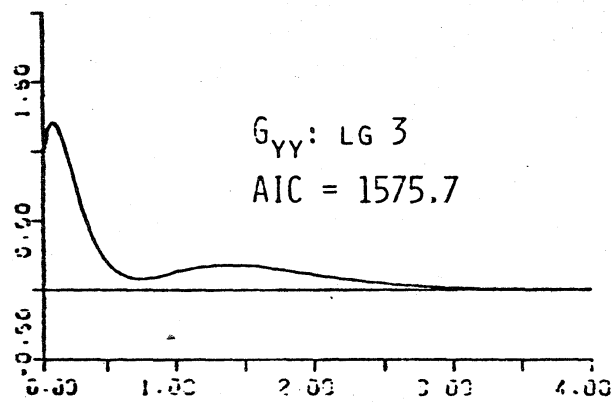
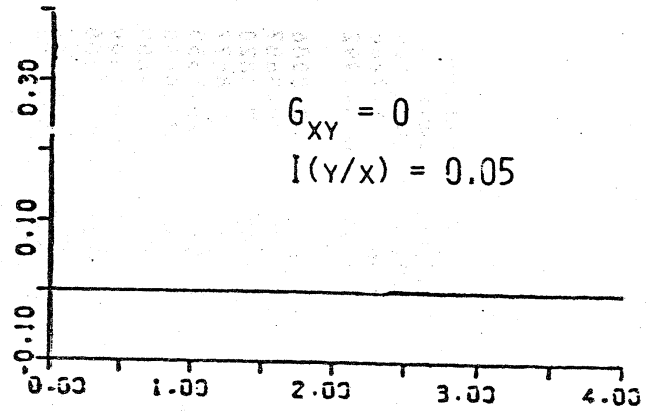
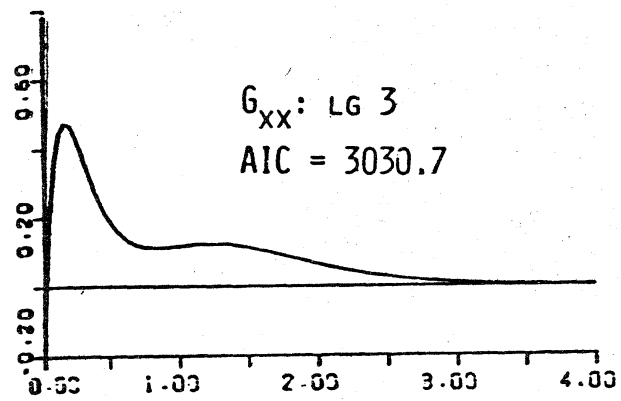
- [1] Akaike, H. (1979). On a point process model which allows efficient likelihood computation, Res. Memo., No. 174, The Inst. Statist. Math., Tokyo.
- [2] Breman, P. M. (1972). A martingale approach to point processes, Memo., No. ERL-M345, Electronics Research Laboratory, Univ. of California, Berkeley.
- [3] Cox, D. R. (1975). Partial likelihood, Biometrika, 62, pp. 269-276.
- [4] Hawkes, A. G. (1971). Spectra of some mutually exciting point processes, Biometrika, 58, pp. 83-90.
- [5] Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes, Ann. Inst. Statist. Math., Vol. 30, No. 2, A, pp. 243-261.
- [6] Ogata, Y. (1979). On Lewis' simulation method for point processes, Res. Memo., No. 147, The Inst. Statist. Math., Tokyo.
- [7] Ozaki, T. & Akaike, H. (1979). A parametric model for Hawkes' self-exciting point processes and its application to the spectrum estimation of point processes, submitted for publication.

1

SIMULATION (THE TRUE MODEL)



⊗ 2 MINIMUM AIC MODEL FOR THE SIMULATION DATA



15

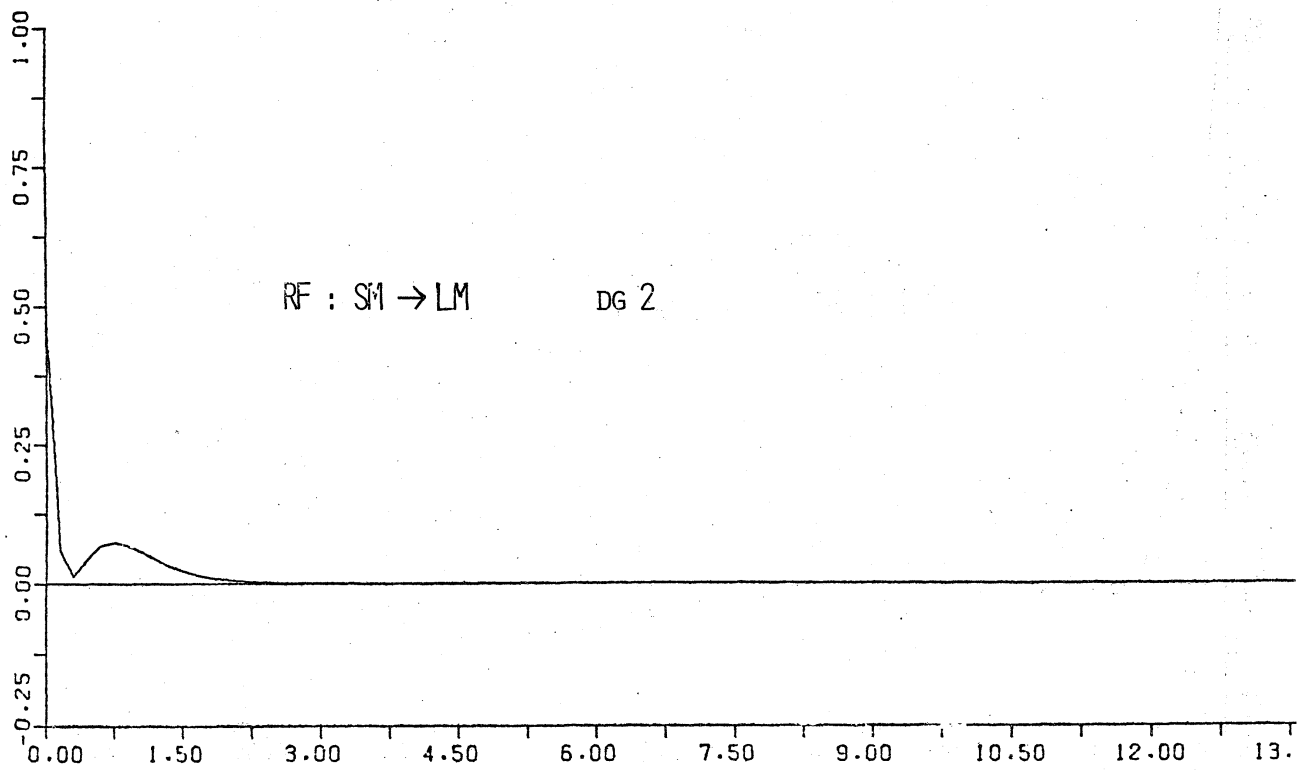
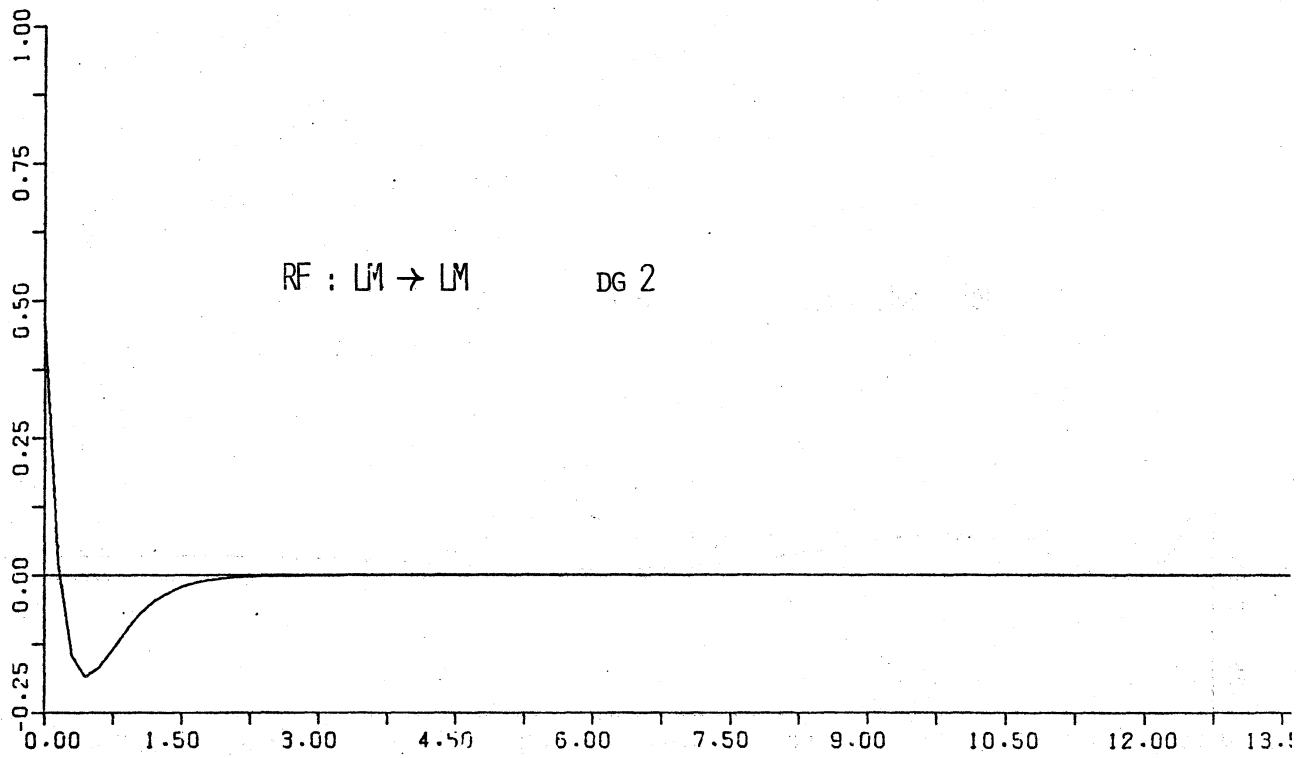
[illegible]

LM : LARGER MAGNITUDE THAN 4.55

S1 : SMALLER MAGNITUDE THAN 4.55

1.0 CORRESPONDS ONE DAY (24H)

3 RESPONSE FUNCTIONS 1



☒ 4 RESPONSE FUNCTIONS 2

