Relatório de Atividade 08 - Métodos de estimação recursivos

Débora Nunes Pinto de Oliveira

Prof. Antonio Marcus, Estimação e Identificação de Sistemas 21.2

28 de setembro de 2021

Esse documento tem por objetivo descrever métodos numéricos iterativos para encontrar o argumento mínimo das funções de custo dos métodos paramétricos de identificação de um processo desconhecido.

Table of Contents

Introdução

Nas atividades anteriores, foram discutidos dois métodos de estimação de parâmetros θ da função de transferência de um processo desconhecido. A primeira abordagem de busca foi descrita a partir da minimização de uma função de custo $V_N(\theta, Z^N)$, para Z^N o conjunto de dados $u(t)$ e $y(t)$ nos instantes anteriores. Em segundo lugar, foi discutida a estimação da função de transferência a partir da correlação f_N entre o vetor de regressores ζ e o erro de predição ε.

Nessa atividade, serão discutidos os métodos numéricos para encontrar a solução das duas abordagens supracitadas. É importante relembrar que, para uma regressão linear, a predição de um processo de saída $y(t)$ é dada por:

$$\hat{y}(t|\theta) = \varphi^T(t)\theta$$

Considerando a função de custo a norma quadrática do erro de predição, o argumento $\hat{\theta}_N^{LS}$ que minimiza $V_N(\theta, Z^N)$ é dado por:

$$\hat{\theta}_N^{LS} = \arg\min V_N(\theta, Z) = \left[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)\right]^{-1} \sum_{t=1}^{N} \varphi(t)y(t) = P^{-1}(N)\sum_{t=1}^{N} \varphi(t)y(t)$$

Entretanto, a matriz $P(N)$ pode ser mal condicionada, isto é, pequenas modificações nos argumentos $\varphi(t)$ resultam em grandes modificações na solução $\hat{\theta}_N^{LS}$. Dessa forma, é adequado empregar métodos numéricos iterativos para encontrar $\hat{\theta}_N^{LS}$ que não envolvam a computação da derivada da função de custo. Na verdade, esses métodos recursivos buscam o mínimo da função de custo a partir de testes de múltiplos valores de θ.

Minimização numérica

A minimização numérica de uma função $V(\theta)$ envolve a atualização da estimativa de θ iterativamente conforme a equação:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \alpha f^{(i)}$$

para $f^{(i)}$ a direção de busca baseada em informações de $V(\theta)$ adquiridas em iterações anteriores, e α uma constante positiva que determina o quanto o resultado de $V(\theta)$ deve decrescer. Os métodos para determinação de $f^{(i)}$ determina a categorização da minimização numérica em três grupos:

1. Métodos usando valores de $V(\theta)$, apenas;
2. Algoritmos de quasi-Newton: Métodos que empregam os valores de $V(\theta)$ e seus gradientes;
3. Algoritmos de Newton: métodos que empregam os valores de $V(\theta)$, seus gradientes e hessianos (derivadas de segundo grau).

Para os algoritmos de Newton, $f^{(i)}$ é determinado como:

$$f^{(i)} = -\left[V''(\hat{\theta}^{(i)})\right]^{-1} V'(\hat{\theta}^{(i)})$$

Considerando $V(\theta, Z) = \frac{1}{N}\sum_{t=1}^{N} \ell\left(\varepsilon(t,\theta), \theta\right)$, o gradiente é determinado como:

$$V'_N(\theta, Z^N) = -\frac{1}{N}\sum_{t=1}^{N} \psi(t,\theta)\ell'_\varepsilon\left(\varepsilon(t,\theta),\theta\right) - \ell'_\theta\left(\varepsilon(t,\theta),\theta\right)$$

para $\psi(t,\theta)$ o gradiente de $\hat{y}(t|\theta)$ em relação a θ. É importante destacar que esse gradiente $V'_N(\theta, Z^N)$ pode ser complexo demais para ser calculado analiticamente. Nesses casos, é apropriado utilizar os métodos do grupo 1 e estimar $\psi(t,\theta)$ por aproximação diferencial.

Especificamente, considerando $\ell(\varepsilon, \theta, t) = \frac{1}{2}\varepsilon(t,\theta)^2$, é possível escrever o gradiente $V'_N(\theta, Z^N)$ como:

$$V'_N(\theta, Z^N) = -\frac{1}{N}\sum_{t=1}^{N} \psi(t,\theta)\varepsilon(t,\theta)$$

Logo, a atualização da estimativa θ é dada por:

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \mu_N^{(i)}\left[R_N^{(i)}\right]^{-1} V'_N(\hat{\theta}^{(i)}, Z^N)$$

para R_N uma matriz que modifica a direção de busca. O passo $\mu_N^{(i)}$ é escolhido de tal forma que $V_N(\hat{\theta}^{(i+1)}, Z^N) < V_N(\hat{\theta}^{(i)}, Z^N)$.

Inicialmente, será considerado o caso do algoritmo quasi-Newton, em que R_N é a matriz identidade. Nessa situação, o método se torna ineficiente próximo ao argumento mínimo $\hat{\theta}_N^{LS}$, uma vez que $V'_N(\hat{\theta}, Z^N) \approx 0$.

Dessa forma, pode-se escolher $R_N^{(i)} = V''_N(\hat{\theta}_N^{(i)}, Z^N)$, para:

$$V''_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \psi(t,\theta)\psi^T(t,\theta) - \frac{1}{N}\sum_{t=1}^{N} \psi'(t,\theta)\varepsilon(t,\theta)$$

Nesse caso, o método de estimativa é um algoritmo de Newton. É importante destacar que pode ser computacionalmente custoso calcular a derivada do gradiente ψ. Sendo assim, considerando que o valor de $\hat{\theta}_N^{LS}$ independe do erro de predição $\varepsilon(\hat{\theta}_N^{LS}, t)$, pode-se aproximar o hessiano como:

$$V''_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \psi(t,\theta)\psi^T(t,\theta)$$

Essa aproximação renomeia o método para Gauss-Newton. Outros métodos, como Levenberg–Marquardt, aproxima R_N como:

$$R_N^{(i)}(\lambda) = V_N'(\widehat{\theta}_N^{(i)}, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \psi(t, \widehat{\theta}_N^{(i)})\psi^T(t, \widehat{\theta}_N^{(i)}) + \lambda I$$

para I a matriz identidade e λ um escalar positivo indicador da convergência das iterações. Para $\lambda = 0$, Levenberg–Marquardt é equivalente ao método de Gauss-Newton. Todavia, para ambos os métodos, a computação do gradiente ψ depende do modelo utilizado para escrever a estimativa $\widehat{y}$ do processo. Para os modelos comuns, é possível padronizar as expressões de ψ analiticamente, como descrito na seção a seguir.

Computando gradientes

Para um modelo de erro na saída, é possível escrever $y(t)$ como:

$$y(t) = \frac{B(q)}{F(q)}u(t) + e(t)$$

para $F(q) = 1 + f_1 q^{-1} + ... + f_{n_f} q^{-n_f}$ e $B(q) = 1 + b_1 q^{-1} + ... + b_{n_b} q^{-n_b}$

Isso significa que há um erro somado diretamente sobre a saída do sistema $w(t) = [B(q)/F(q)]u(t)$, isto é:

$$w(t) + f_1 w(t-1) + ... + f_{n_f} w(t - n_f) = b_1 u(t-1) + ... + b_{n_b} u(t - n_b)$$

Nessa planta, $w(t)$ não é diretamente observado, mas obtido a partir da expressão acima para cada aproximação dos coeficientes f_i e b_i. Dessa forma, $w(t, \theta)$ é alterado a cada iteração do algoritmo de predição.

Seja o vetor θ é composto por $\theta = \left[b_1\ b_2\ ...\ b_{n_b}\ f_1\ f_2\ ... f_{n_f}\right]$, logo, o vetor $\varphi(t, \theta)$ é dado por:

$$\varphi(t, \theta) = \left[u(t-1)\ ...\ u(t-n_b)\ w(t-1, \theta)\ ...\ w(t-n_c, \theta)\right].$$

É evidente que esse modelo então é pseudolinear. O preditor, nesse caso, é dado por:

$$\widehat{y}(t|\theta) = \frac{B(q)}{F(q)}u(t)$$

Diferenciando o preditor em relação a b_k e f_k, tem-se que:

$$\frac{\delta}{\delta b_k}\widehat{y}(t|\theta) = -\frac{1}{F(q)}u(t-k) \text{ e } \frac{\delta}{\delta f_k}\widehat{y}(t|\theta) = -\frac{1}{F(q)F(q)}u(t-k) = -\frac{1}{F(q)}w(t-k, \theta)$$

Sendo assim, é possível reescrever que $F(q)\psi(t\theta) = \varphi(t, \theta)$. O gradiente é então obtido filtrando o vetor de regressão por um filtro $1/F(q)$. Para encontrar uma estimativa $\widehat{\theta}$ para um regressor não linear, é recomendado empregar o método dos mínimos quadrados em duas ou mais etapas, conforme descrito na seção a seguir.

Métodos multi-etapas

Uma forma de calcular o mínimo da função de custo para um problema não linear é reformular como um conjunto de regressões lineares. Esses métodos geralmente envolvem dois ou mais estágios de mínimos quadrados.

Por exemplo, considere que a função de custo a ser minimizada é a função de correlação:

$$f_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \zeta(t, \theta)\varepsilon(t, \theta)$$

para $\zeta(t, \theta)$ o vetor de variáveis instrumentais ou qualquer função filtrada de $u(t)$. Nesse estudo, será considerado que $\zeta(t) = \psi(t, \theta)$. A estimativa $\widehat{\theta}_N$ é então computada, por mínimos quadrados, como:

$$\hat{\theta}_N^{(i+1)} = \left[\sum_{t=1}^{N} \zeta(t, \hat{\theta}_N^{(i)}) \varphi^T(t, \hat{\theta}_N^{(i)})\right]^{-1} \sum_{t=1}^{N} \zeta(t, \hat{\theta}_N^{(i)}) y(t)$$

Como esse método alterna entre computar $\hat{\theta}$ e os vetores ζ e φ, ele é nomeado *bootstrap*.

Desenvolvimento

Nessa atividade, serão solucionados dois exercícios do livro texto: 10G.2 e 11T.1. Ambas as atividades propõem o uso do método multi-etapas.

Primeiramente, será avaliada a aplicação de um modelo de erro na saída na estimativa de parâmetros via método multi-etapas. Em seguida, será discutida a implementação recursiva da atualização das estimativas pelo método de Newton-Raphson, e comparada essa implementação com o algoritmo recursivo do método multi-etapas.

Questão 10G.2

Steiglitz e McBride sugeriram o seguinte método para a identificação de um sistema linear sujeito a erros de medição modelados conforme um ruído branco:

Considere um sistema de erro na saída, modelado conforme:

$$y(t) = \frac{B(q)}{F(q)} u(t) + e(t)$$

Primeiramente, aplica-se o método de mínimos quadrados para encontrar os polinômios $\hat{B}_N(q)$ e $\hat{F}_N(q)$ sobre o modelo ARX:

$$F(q) y(t) = B(q) u(t) + e(t)$$

Em seguida, filtra-se os dados pelo filtro $1/\hat{F}_N(q)$, de tal forma que:

$$y_F(t) = \frac{1}{\hat{F}_N(q)} y(t) \text{ e } u_F(t) = \frac{1}{\hat{F}_N(q)} u(t)$$

Dessa forma, aplica-se novamente o método dos mínimos quadrados sobre o modelo do ARX:

$$F(q) y_F(t) = B(q) u_F(t) + e(t)$$

A filtragem e a segunda estimativa por mínimos quadrados deve então ser repetida até que as estimativas de $\hat{F}_N$ e $\hat{B}_N$ se adequem à:

$$-\beta \leq \hat{F}_N^{(i)} - \hat{F}_N^{(i-1)} \leq \beta \text{ e } -\beta \leq \hat{B}_N^{(i)} - \hat{B}_N^{(i-1)} \leq \beta$$

para β um escalar positivo que indique a tolerância do algoritmo

Conforme citado na seção de métodos multi-etapas, esse método pode ser interpretado como uma filtragem de $u(t)$ por um filtro dependente de θ. Sendo assim, considerando que as estimativas do polinômio $F(q)$ pertencem ao vetor θ, é possível considerar que $\zeta(t, \theta)$ composto pelos dados $u_F(t)$ equivalem a filtrar os dados $u(t)$, a cada passo de estimativa, pelo filtro $1/\hat{F}_N^{(i)}$. Conforme a notação da Atividade 7, o erro de predição $\varepsilon_F(t, \theta) = L(q)\varepsilon(t, \theta)$, para $L(q) = F(q)$. Nesse caso, como $F(q)\psi(t\theta) = \zeta(t, \theta)$, é evidente que esse algoritmo computa o gradiente e, portanto, é denominado quasi-Newton.

Como $\varepsilon_F(t, \theta) = e_0(t)$ (para $e_0(t)$ um ruído branco) independe de valores passados de $u(t)$, e $\zeta(t, \theta)$ depende unicamente de valores passados de $u(t)$, é evidente que a correlação entre $\zeta(t, \theta)$ e $\varepsilon_F(t, \theta)$ é nula. Sendo assim, a consistência do algoritmo de mínimos quadrados depende unicamente se a matriz $\zeta(t, \theta)\varphi_F^T(t)$ é não singular, para $\varphi_F(t) = F(q)\varphi(t)$.

Por sua vez, a variância assintótica P_θ de $f_N = 1/N \sum_{t=1}^{N} \zeta(t, \theta)\varepsilon(t, \theta)$ é dada por:

$$P_\theta = [\overline{f}'_N(\theta_0)]^{-1} Q [\overline{f}'_N(\theta_0)]^{-T}$$

para $Q = \lim_{N\to\infty} E \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} \zeta(t,\theta_0)\varepsilon_F(t,\theta_0)\varepsilon_F^T(t,\theta_0)\zeta^T(t,\theta_0)$ e $\overline{f}'_N(\theta) = \overline{E}\zeta'_\theta(t,\theta)\varepsilon_F(t,\theta) - \overline{E}\zeta(t,\theta)[L(q)\psi(t,\theta)]^T$.

Como para $\varepsilon_F(t,\theta) = e_0(t)$ a correlação nula entre $\zeta(t,\theta)$ e $\varepsilon_F(t,\theta)$, para $e_0(t)$ um ruído branco de variância λ_0, então:

$$Q = \lambda_0 \overline{E}\zeta(t,\theta_0)\zeta^T(t,\theta_0) \text{ e } P_\theta = \lambda_0 \left[\overline{E}\zeta(t,\theta_0)\left[L(q)\psi(t,\theta)\right]^T\right]^{-1} Q[\overline{E}\zeta(t,\theta_0)\left[\overline{E}L(q)\psi(t,\theta)\left[\zeta(t,\theta_0)\right]^T\right]^{-1}$$

Questão 11T.2

Para o método de Newton-Raphson, a aproximação da estimativa $\widehat{\theta}$ é dada por:

$$\widehat{\theta}^{(i+1)} = \widehat{\theta}^{(i)} - \mu_N^{(i)}\left[f'_N(\widehat{\theta}^{(i)}, Z^N)\right]^{-1} f_N(\widehat{\theta}^{(i)}, Z^N)$$

para $f_N = 1/N \sum_{t=1}^{N} \zeta(t,\theta)\alpha(\varepsilon(t,\theta))$. Considerando $\mu_N^{(i)} = 1$ e $f'_N(\widehat{\theta}^{(i)}, Z^N) = R_N^{(i)}$, então é possível reescrever a equação acima como:

$$\widehat{\theta}^{(i+1)} = \widehat{\theta}^{(i)} - [R_N^{(i)}]^{-1} f_N(\widehat{\theta}^{(i)}, Z^N)$$

Calculando $R_N^{(i)}$ analiticamente, tem-se:

$$R_N^{(i)} = \frac{1}{N}\sum_{t=1}^{N}\left[\zeta'_\theta(t,\theta)\alpha(\varepsilon(t,\theta)) + \zeta(t,\theta)\alpha'_\varepsilon(\varepsilon(t,\theta))\psi(t,\theta)\right]$$

Além disso, pode-se calcular $f_N\left(\widehat{\theta}^{(N)}, Z^N\right)$ como:

$$f_N\left(\widehat{\theta}^{(N)}, Z^N\right) = f_{N-1}\left(\widehat{\theta}^{(N-1)}, Z^{N-1}\right) + \frac{1}{N}\left[\zeta(N, \widehat{\theta}^{(N-1)})\alpha(\varepsilon(N, \widehat{\theta}^{(N-1)})) - f_{N-1}\left(\widehat{\theta}^{(N-1)}, Z^{N-1}\right)\right]$$

Assumindo que a estimativa anterior $\widehat{\theta}^{(N-1)}$ minimizou $f_{N-1}\left(\widehat{\theta}^{(N-1)}, Z^{N-1}\right)$ de tal forma que $f_{N-1}\left(\widehat{\theta}^{(N-1)}, Z^{N-1}\right) = 0$, substituindo em $f_N\left(\widehat{\theta}^{(N)}, Z^N\right)$, obtém-se:

$$f_N\left(\widehat{\theta}^{(N)}, Z^N\right) = \frac{1}{N}\zeta(N,\theta)\alpha(\varepsilon(N,\theta))$$

Logo, a estimativa $\widehat{\theta}$ é dada por (para $\mu_N = 1$):

$$\widehat{\theta}^{(i+1)} = \widehat{\theta}^{(i)} - \frac{1}{N}[R_N]^{-1}\zeta(N,\theta)\alpha(\varepsilon(N,\theta))$$

Analogamente, a recursividade sobre $R_N^{(i)}$ é dada como:

$$R_N^{(i)} = R_N^{(i-1)} + \frac{1}{N}\left[\zeta'_\theta(N,\theta)\alpha(\varepsilon(N, \widehat{\theta}^{(N-1)})) + \zeta(N, \widehat{\theta}^{(N-1)})\alpha'_\varepsilon(\varepsilon(N, \widehat{\theta}^{(N-1)}))\psi(N, \widehat{\theta}^{(N-1)}) - R_N^{(i-1)}\right]$$

Esse algoritmo de Newton-Raphson recursivo e o método em multi-etapas diferem apenas na definição da matriz $R_N^{(i)}$. Para o primeiro, $R_N^{(i)} = f'_N(\widehat{\theta}^{(i)}, Z^N)$. Já para o segundo, $R_N^{(i)} = \sum_{t=1}^{N} \zeta(t, \widehat{\theta}_N^{(i)})\varphi^T(t, \widehat{\theta}_N^{(i)})$, ou seja, a função de correlação entre o vetor de regressores filtrado e original.