

Оценивание по конечному числу наблюдений

Олег Николаевич Граничин

Санкт-Петербургский государственный университет,
математико-механический факультет

27 февраля 2013

Если имеется выборочная последовательность $\varphi_1, \varphi_2, \dots, \varphi_N$ реализаций случайной величины η , то в рамках линейной регрессионной модели реализации y_1, y_2, \dots, y_N случайной величины ζ удобно представить в виде

$$y_n = \varphi_n^T \mathbf{x}^* + v_n, \quad n = 1, 2, \dots, N,$$

где \mathbf{x}^* — вектор коэффициентов. В этом представлении невязки v_n , $n = 1, 2, \dots, N$, интерпретируются как ошибки наблюдения. Обозначим $\mathbf{y} = \mathbf{y}_N = (y_1, y_2, \dots, y_N)^T$ — наблюдаемый в момент времени N вектор, являющийся функцией входных воздействий, помех в канале измерения и некоторого векторного параметра \mathbf{x} .

Требуется по значению вектора \mathbf{Y} получить хорошую оценку-вектор $\hat{\mathbf{x}} = \hat{\mathbf{x}}_N$ вектора \mathbf{x} или множество $\hat{X} = \hat{X}_N$, содержащее вектор θ .

- Оценка $\hat{\mathbf{x}}$ называется *линейной*, если она имеет вид $\hat{\mathbf{x}} = \Gamma \mathbf{y}$ с некоторой матрицей коэффициентов Γ .
- Оценка $\hat{\mathbf{x}}$ называется *несмещенной*, если $E\{\hat{\mathbf{x}}\} = \mathbf{x}^*$.
- Последовательность оценок $\{\hat{\mathbf{x}}_N\}_{N=1}^{\infty}$ называется *состоятельной*, если для любого $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\{\|\hat{\mathbf{x}}_N - \mathbf{x}^*\|^2 > \varepsilon\} = 0,$$

- и называется *сильносостоятельной*, если с вероятностью единица

$$\lim_{N \rightarrow \infty} \hat{\mathbf{x}}_N = \mathbf{x}^*.$$

- Множество \hat{X}_N называется *гарантированным*, если $\mathbf{x}^* \in \hat{X}_N$.
- Множество \hat{X}_N называется *доверительным с уровнем конфиденциальности p* , если $\mathbf{x}^* \in \hat{X}_N$ с вероятностью p .

Для характеристики качества оценки $\hat{\mathbf{x}}$ используется теоретически предсказываемый выбранной моделью выходной сигнал \mathbf{z} , т. е. выход принятой модели, который зависит от $\hat{\mathbf{x}}$. Эта зависимость, вообще говоря, может быть выбрана различными способами. Простейшей является линейная:

$$\mathbf{z} = \Phi^T \hat{\mathbf{x}},$$

в которой матрица Φ состоит из векторов $\varphi_1, \varphi_2, \dots, \varphi_N$

$$\Phi = (\varphi_1, \varphi_2, \dots, \varphi_N).$$

Ошибка оценивания естественно определить как $\mathbf{v} = \mathbf{y} - \mathbf{z}$. Оценки $\hat{\mathbf{x}}$, минимизирующие функционал качества (функцию потерь)

$$f_N(\mathbf{x}) = \|\mathbf{V}\|^2 = \sum_{n=1}^N \|y_n - \varphi_n^T \mathbf{x}\|^2 = \sum_{n=1}^N \|v_n\|^2,$$

называются *оценками метода наименьших квадратов (МНК)*.

Функционал $f_N(\mathbf{x})$ зависит от выборочных значений реализаций случайных величин и называется *эмпирическим*.

В ситуации использования бесконечной выборки последовательность величин $f_N(\mathbf{x})/N$ может в том или ином вероятностном смысле стремиться, в силу закона больших чисел, при $N \rightarrow \infty$ к значению квадратичного функционала, определяющего линейную регрессию. В частности, если предположить независимость ошибок наблюдения и ограниченность их вторых моментов, то сходимость будет с вероятностью единица.

Несложно убедиться, что оценки МНК должны удовлетворять системе уравнений

$$\Phi\Phi^T\hat{\mathbf{x}} = \Phi\mathbf{y}.$$

Если матрица $\Phi\Phi^T$ невырождена, то эта система имеет единственное решение

$$\hat{\mathbf{x}} = (\Phi\Phi^T)^{-1}\Phi\mathbf{y}.$$

Можно рассмотреть более общее правило выбора оценки. Пусть R — некоторая симметричная неотрицательная матрица *весовых коэффициентов*. Введём функционал качества по правилу

$$f_N(\mathbf{x}) = \mathbf{v}^T R \mathbf{v}.$$

Для линейной модели с матрицей Φ при невырожденной (имеющей обратную) матрице $\Phi R \Phi^T$ оптимальная оценка имеет вид

$$\hat{\mathbf{x}} = (\Phi R \Phi^T)^{-1} \Phi R \mathbf{y} = \Gamma \mathbf{y}, \quad \Gamma = (\Phi R \Phi^T)^{-1} \Phi R,$$

и называется *обобщенной оценкой МНК*.

При этом, если принять, что

$$\mathbf{y} = \Phi^T \mathbf{x}^* + \mathbf{v},$$

то векторы \mathbf{x}^* и $\hat{\mathbf{x}}$ связаны соотношением

$$\mathbf{x}^* = \hat{\mathbf{x}} - \Gamma \mathbf{v}.$$

Последнее соотношение справедливо при любой природе ошибки оценивания. Если предположить, что условное математическое ожидание $E\{\mathbf{v}|\Phi\} = 0$, то полученная оценка является несмещенной и

$$E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2|\Phi\} = \text{Tr}[\Gamma B_{\mathbf{v}} \Gamma^T], \quad \text{cov}\{\hat{\mathbf{x}} \hat{\mathbf{x}}^T|\Phi\} = \Gamma B_{\mathbf{v}} \Gamma^T,$$

где $B_{\mathbf{v}} = E\{\mathbf{v} \mathbf{v}^T|\Phi\}$.

Если матрица $B_{\mathbf{v}}$ имеет обратную, то можно выбрать $R = B_{\mathbf{v}}^{-1}$.

Получаемые оценки

$$\hat{\mathbf{x}} = (\Phi B_{\mathbf{v}}^{-1} \Phi^T)^{-1} \Phi B_{\mathbf{v}}^{-1} \mathbf{Y}$$

называются *марковскими*. Условная ковариация марковской оценки при $E\{\mathbf{v}|\Phi\} = 0$ равна

$$\text{cov}(\hat{\mathbf{x}}^T \hat{\mathbf{x}} | \Phi) = (\Phi B_{\mathbf{v}}^{-1} \Phi^T)^{-1}.$$

Theorem

(Гаусса–Маркова)

Если $E\{\mathbf{v}|\Phi\} = 0$ и матрица $B_{\mathbf{v}}$ обратима,

то среди всех линейных оценок вида $\tilde{\mathbf{x}} = (\Phi R \Phi^T)^{-1} \Phi R \mathbf{y}$ условная ковариация для марковской оценки $\hat{\mathbf{x}}$ минимальна в том смысле, что

$$\text{cov}(\hat{\mathbf{x}}^T \hat{\mathbf{x}} | \Phi) \leq \text{cov}(\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} | \Phi).$$

Доказательство теоремы Гаусса–Маркова

Воспользуемся матричным неравенством для невырожденной матрицы $A^T A$

$$B^T B \geq B^T A (A^T A)^{-1} A^T B,$$

которое выполнено для произвольной матрицы B соответствующей размерности.

Подставив в него $A = B_{\mathbf{v}}^{-1/2} \Phi^T$, $B^T = \Gamma B_{\mathbf{v}}^{1/2}$, получаем заключение теоремы:

$$\begin{aligned} \text{cov}(\tilde{\mathbf{x}}^T \tilde{\mathbf{x}} | \Phi) &= \Gamma B_{\mathbf{v}} \Gamma \geq \Gamma B_{\mathbf{v}}^{1/2} B_{\mathbf{v}}^{-1/2} \Phi^T (\Phi B_{\mathbf{v}}^{-1/2} B_{\mathbf{v}}^{-1/2} \Phi^T)^{-1} \Phi B_{\mathbf{v}}^{-1/2} B_{\mathbf{v}}^{1/2} \Gamma^T = \\ &= \Gamma \Phi^T (\Phi B_{\mathbf{v}}^{-1} \Phi^T)^{-1} \Phi \Gamma^T = \text{cov}(\hat{\mathbf{x}}^T \hat{\mathbf{x}} | \Phi). \end{aligned}$$

Пример

Рассмотрим задачу об оценивании неизвестного коэффициента усиления x^* известного скалярного сигнала $\{\varphi_n\}_{n=1}^N$, наблюдаемого на фоне помех. Пусть полезный сигнал, помехи и наблюдаемые величины связаны системой уравнений:

$$y_n = \varphi_n x^* + v_n, \quad n = 1, 2, \dots, N,$$

в которой y_1, y_2, \dots, y_N — наблюдения и v_1, v_2, \dots, v_N — помехи. Предположим, что помехи $\{v_n\}_{n=1}^N$ представляют собой реализацию стохастически независимых одинаково распределенных центрированных случайных величин с дисперсией σ_v^2 :

$$E\{v_n\} = 0, \quad E\{v_n^2\} = \sigma_v^2 > 0.$$

Введя соответствующие обозначения $\mathbf{y}, \Phi, \mathbf{v}$ получаем

$$\mathbf{y} = \Phi^T x^* + \mathbf{v}.$$

Если $\sum_{n=1}^N \varphi_n^2 > 0$, то оценка МНК имеет вид

$$\hat{x} = \frac{\sum_{n=1}^N \varphi_n y_n}{\sum_{n=1}^N \varphi_n^2}.$$

Марковские оценки в данном случае совпадают с оценками МНК, что является следствием некоррелированности и одинаковой распределённости случайных величин $\{v_n\}_{n=1}^N$. В той ситуации, когда полезный сигнал $\{\varphi_n\}_{n=1}^N$ и помехи $\{v_n\}_{n=1}^N$ независимы, дисперсия оценки равна

$$\sigma_N^2 = \frac{\sigma_v^2}{\sum_{n=1}^N \varphi_n^2}.$$

Пусть $N \rightarrow \infty$. Если

$E\{\varphi_n\} = M_\varphi$, $E\{(\varphi_n - M_\varphi)^2\} = \sigma_\varphi^2 > 0$, $E\{\varphi_n^4\} < \infty$, $E\{\varphi_n v_n\} = 0$, то, в силу усиленного закона больших чисел $\frac{1}{N} \sum_{n=1}^N \varphi_n^2 \rightarrow \sigma_\varphi^2$ с вероятностью единица. Следовательно, $\sigma_N^2 \rightarrow 0$, т. е. оценки МНК сильносостоятельные.

Модели авторегрессии и скользящего среднего

Кроме линейной регрессии часто изучают более общие модели:
скользящего среднего

$$y_n = \sum_{i=0}^p \varphi_{n-i}^T b_i^* + v_n, \quad n = 1, 2, \dots, N;$$

авторегрессии

$$\sum_{i=0}^p a_i^* y_{n-i} = \varphi_n^T b_0^* + v_n, \quad n = 1, 2, \dots, N$$

и авторегрессии скользящего среднего

$$\sum_{i=0}^p a_i^* y_{n-i} = \sum_{i=0}^p \varphi_{n-i}^T b_i^* + v_n, \quad n = 1, 2, \dots, N.$$

Модель авторегрессии возникает при попытках определения линейного уравнения, связывающего выборочные реализации случайной величины ζ . Модель авторегрессии скользящего среднего используется для описания поведения линейных динамических объектов управления.

Compressive Sensing

Information processing and control decision making may be schematized as follows

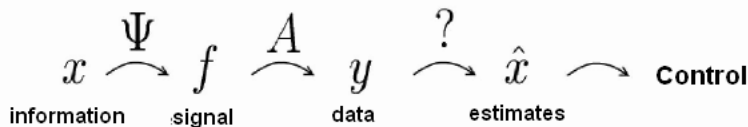


Рис.: Diagram of data acquisition and processing.

We assume that

- the information x about an important for the researcher variable phenomenon (or control plant, or some element of the environment)
- mediated by some forms of communication Ψ ,
- manifests itself via the signal f .

The modern information theory originated from the famous *Kotel'nikov–Nyquist–Shannon theorem* stating that if an analog signal $f : \mathbb{R} \rightarrow \mathbb{R}$ from $L_2(\mathbb{R})$ has a limited spectrum, then it can be restored uniquely without losses from its discrete readings taken at a frequency greater than the doubled maximal frequency of the spectrum. This theorem in many cases enables one to identify the analog signal $f(t)$ with the corresponding set of discrete values f_n , $n = 1, 2, \dots, N$.

The volumes of information processed have increased dramatically in the XXI century. To a large measure, this is due to the mass transition to processing the flows of two-dimensional (2-D) and three-dimensional (3-D) data. Complexity of the traditional methods of signal quantization grows exponentially with dimensionality. Quantization of the 1-D signals for $N = 10^3$ readings corresponds to $N = 10^6$ for the 2-D case and $N = 10^9$ for 3-D, which is extremely high. In the modern applications for the digital photo and video cameras, the traditional requirement on the desired measurement frequency (speed, Nyquist rate) is so high that too much data must be substantially compressed before being stored or transmitted. In other applications including the display systems (medical scanners and radars) and high-speed analog-digit converters, the increase in the measurement frequency proves to be too costly.

Sparse (compressive) Signals

Let f be $N \times 1$ column vector $\in \mathbb{R}^N$. Any signal in \mathbb{R}^N may be expanded in some basis of the $N \times 1$ -dimensional vectors $\{\psi_j\}_{j=1}^N$

$$f = \Psi x = \sum_{j=1}^N x_j \psi_j, \quad (1)$$

where $\Psi = \text{col}(\psi_1, \psi_2, \dots, \psi_N)$, x is an $N \times 1$ column vector of the weight coefficients $x_j = \langle x, \psi_j \rangle = \psi_j^T x$.

f is called the representation in the time (or space) domain, and x is one in the spectral or transformed Ψ -domain.

The signal f is referred to as *s-sparse* if it is a linear combination only of s basic vectors, that is, only s components x_j in (1) are other than zero and the rest ($N - s$) of them are zero.

Of interest is the case where $s \ll N$.

3-sparse vector $x = \text{col}(0, \dots, 0, 1, 0, \dots, 7, 0, \dots, 0, 3, 0, \dots)$

One of the following orthogonal transforms is usually employed in image processing: discrete Fourier transform (DFT), discrete cosine transform (DCT) or discrete wavelet-transform (DWT).

Unfortunately, the transforming coding where a full sample is first obtained and then compression is performed has three inherent disadvantages:

- the initial size of the sample N may be extremely large even if the size s of the resulting assembly of components is small.
- all N transformation coefficients $\{x_j\}$ must be calculated even though their majority, except for s coefficients, will be rejected.
- the positions of s large coefficients must be additionally coded.

Compressive sensing

- $m \ll N$
- compressive (sparse) signals

Candes E, Romberg J, Tao T (2006) Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information // IEEE Trans Inform Theor 52(2):489–509

Donoho D (2006) Compressed sensing // IEEE Trans Inform Theory 52(4):1289–1306

The problem of compressive sensing lies in designing

- How can we choose a universal matrix A ?
(a universal measurement matrix A such that the essential information about any s -sparse (or compressed) signal is not damaged at reducing the dimensionality from $f \in \mathbb{R}^N$ to $y \in \mathbb{R}^m$)
- What is a signal recovery algorithm?
(a reconstruction algorithm restoring x — and, consequently, f —only from $m \sim s$ measurements)

- Restricted Isometric Property

$$\sqrt{1-\delta} \leq \frac{\|\Phi z\|_2}{\|z\|_2} \leq \sqrt{1+\delta} \quad (2)$$

- Incoherence

$$\mu(A, \Psi) = \sqrt{N} \max_{i,j} \frac{|\langle a_i, \psi_j \rangle|}{\|a_i\|_2},$$

Universal Random Matrix A ?

The random $m \times N$ measurement matrix A with independent identically distributed (i.i.d.) elements $a[i,j]$

- Gaussian with zero means and variance $1/m$
- Bernoulli

$$P(a[i,j] = \pm 1/\sqrt{m}) = \frac{1}{2}$$

or a similar distribution

$$a[i,j] = \begin{cases} +\sqrt{3/m} & \text{with probability } \frac{1}{6} \\ 0 & \text{с вероятностью } \frac{2}{3} \\ -\sqrt{3/m} & \text{with probability } \frac{1}{6} \end{cases}$$

- ...

$$m \geq c_1 s \log(N/s), \quad (3)$$

Optimization based on the ℓ_1 -norm

$$\hat{x} = \arg \min \|x'\|_1 : \Phi x' = y,$$

enables one to reconstruct precisely and with high probability the s -sparse signals and approximate well the compressed signals using only

$$m \geq c_1 s \log(N/s)$$

i.i.d. random measurements.

Geometrical Interpretation

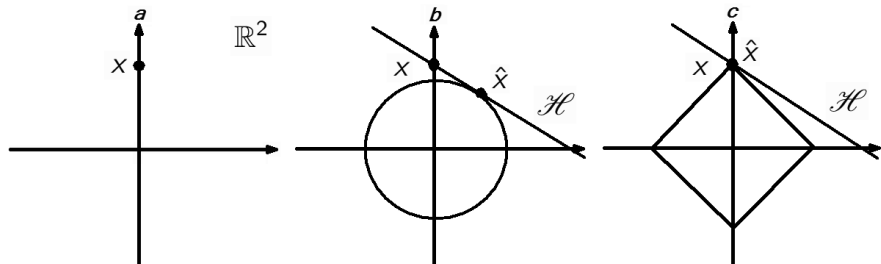


Рис.: (a) The space of 1-sparse vectors in \mathbb{R}^2 consists of two lines having one non-zero coordinate (axes); (b) ℓ_2 -minimization determines the non-sparse vector \hat{x} distant from x ; (c) ℓ_1 -minimization determines the sparse point \hat{x} of contact of the ℓ_1 -sphere with the hyperplane \mathcal{H} coinciding with x with high probability

Example

Let the signal $f(t)$, $t \in [0,1)$ be a sum of five sinusoidal functions and noise:

$$f(t) = 10.7 \sin(2\pi \cdot 50t) +$$

$$+ 20 \sin(2\pi \cdot 120t) + 31.5 \sin(2\pi \cdot 200t) +$$

$$+ 23 \sin(2\pi \cdot 300t) + 25 \sin(2\pi \cdot 450t) + \text{error}$$

$$m \approx 4 \cdot 5 \cdot \log_2(1000/5) (= 153).$$

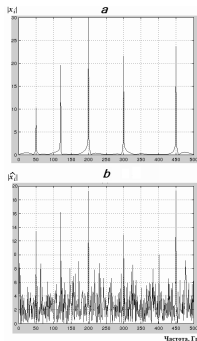
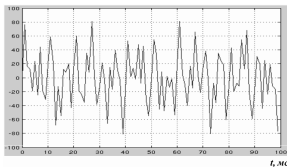


Рис.: Determination of frequencies and amplitudes of the 5-sparse signal. (a) The original and (b) restored signals in the spectral domain.

Рис.: 5-sparse noisy signal.

A single-pixel camera

<http://dsp.rice.edu/cscamera>

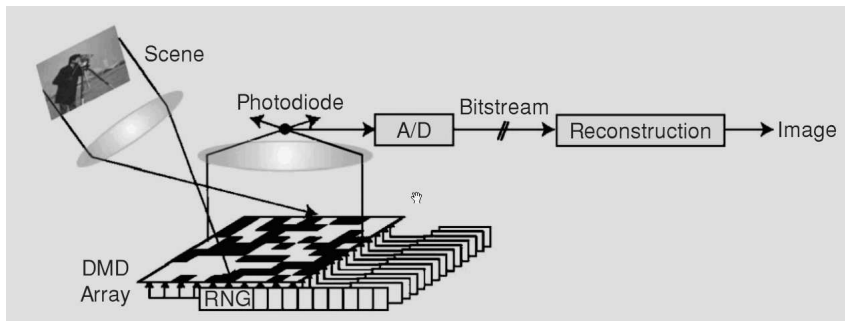


Рис.: One pixel camera.

The light waves f emerging from the photograph pass through the first lens and get into the array of two-position micromirrors reflecting light either in the direction of the second lens or aside.

Amelin K., and Granichin O. Randomized controls for linear plants and confidence regions for parameters under external arbitrary noise // In: Proc. of 2012 American Control Conference, Montreal, Canada, June 27-June 29, 2012. P. 851–856.

Let's consider the dynamical system:

$$A_*(z^{-1})y_t = B_*(z^{-1})u_t + v_t, \quad t = 1, 2, \dots, N, \quad (4)$$

where

$$A_*(\lambda) = 1 + a_*^{(1)}\lambda + \dots + a_*^{(n_a)}\lambda^{n_a}, \quad B_*(\lambda) = b_*^{(l)}\lambda^l + b_*^{(l+1)}\lambda^{l+1} + \dots + b_*^{(n_b)}$$

$\tau_* = \text{col}(a_*^{(1)}, \dots, a_*^{(n_a)}, b_*^{(l)}, \dots, b_*^{(n_b)})$ is the vector of parameters some of which are unknown.

Control Strategy Randomization

$$\text{Goal: } \overline{\lim}_{t \rightarrow \infty} |y_t| \rightarrow \min, \sup_t |y_t| + |u_t| < \infty$$

(Granichin, Fomin, ARC, 1986)

$$s \in \mathbb{N}, N = s \cdot N_{\Delta}.$$

$$u_{sn+i-l} = \begin{cases} \Delta_n + \bar{u}_{sn-l}, & i = 0, \\ \bar{u}_{sn+i-l}, & i = 1, 2, \dots, s-1, \end{cases} \quad n = 0, \dots, N_{\Delta} - 1,$$

$$\bar{u}_t = \mathcal{U}_t(y_t, y_{t-1}, \dots, \bar{u}_{t-1}, \dots), \quad t \geq 0, \quad \bar{u}_{-k} = 0, \quad k > 0.$$

Δ_n is a measurable random trial perturbation

$$\text{For example, } \Delta_n = \begin{cases} +1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}. \end{cases}$$

A1. The user can choose Δ_n and this choice does not affect to the external noise $v_{sn}, \dots, v_{s(n+1)-1}$.

Reparametrization

We can rewrite the model (4) as a linear regression

$$y_{sn+k-1} = \Delta_n \theta_\star^{(k)} + \bar{v}_{sn+k-1} = \Delta_n \theta_\star^{(k)} + \sum_{i=0}^{k-1} \theta_\star^{(k-i)} \bar{u}_{sn-l+i} + \dots, \quad (5)$$

with inputs Δ_n and regressors $\theta_\star^{(k)}$, $n = 1, \dots, N_\Delta$, $k = 1, \dots, s$.

$$\theta_\star = \theta(\tau_\star), \quad \theta(\tau) = \mathbb{A}^{-1}(\tau) \mathbb{B}(\tau), \quad (6)$$

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ a_\star^{(1)} & 1 & \dots & 0 & 0 \\ a_\star^{(2)} & a_\star^{(1)} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & a_\star^{(n_a)} & \dots & a_\star^{(1)} & 1 \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} b_\star^{(l)} \\ \vdots \\ b_\star^{(n_b)} \\ \vdots \\ 0 \end{pmatrix}.$$

(Granichin, Fomin, ARC, 1986)

Note, that Δ_n and \bar{v}_{sn+k-1} are independent if the user can choose Δ_n .

How can we choose s ?

$$\mathbf{A2.} \quad s : \exists \tau(\theta) = \theta^{-1}(\tau)$$

$s = n_a + n_b$ if polynomials $A_*(\lambda)$ and $B_*(\lambda)$ are mutually prime

Example Consider the second-order plant

$$y_t + a_*^{(1)} y_{t-1} + y_{t-2} = b_*^{(1)} u_{t-1} + 1.6 u_{t-2} + v_t, \quad (7)$$

$t = 1, 2, \dots, N$, with unknown coefficients $a_*^{(1)}$ and $b_*^{(1)} \neq 0$. Denote

$$\tau_* = \text{col}(a_*^{(1)}, b_*^{(1)}).$$

Let $s = 2$, vector θ_* of the inverse function $\tau(\theta)$ be

$$\theta_* = \begin{pmatrix} b_*^{(1)} \\ 1.6 - a_*^{(1)} b_*^{(1)} \end{pmatrix} \in \mathbb{R}^2, \quad \tau(\theta) = \begin{pmatrix} \frac{1.6 - \theta^{(2)}}{\theta^{(1)}} \\ \theta^{(1)} \end{pmatrix}.$$

Algorithm 1. Stochastic Approximation With Randomized Inputs

$$\widehat{\theta}_n^{(i)} = \widehat{\theta}_{n-1}^{(i)} - \frac{1}{n} \Delta_n (u_{sn-l} \widehat{\theta}_{n-1}^{(i)} - y_{sn+i-1}), \quad i \in [1..s]. \quad (8)$$

Theorem

If the conditions **A1–A2** and

$$2\sigma_{\Delta}^2 > 1, \quad E\{v_t^2\} \leq \sigma_v^2, \quad \sum_{n=1}^{\infty} \frac{\beta_n^{-1}}{n} = \infty, \quad \sum_{n=1}^{\infty} \frac{\beta_n^{-2}}{n^2} < \infty$$

are satisfied **then** for an arbitrary initial vector $\widehat{\theta}_0 \in \mathbb{R}^s$ the algorithm (8) ensures the estimates $\{\widehat{\theta}_n\}$ such that for an arbitrary $\rho > 0$ the following limit bounds are valid in the mean square sense:

$$E\{|\widehat{\theta}_n^{(i)} - \theta_{\star}^{(i)}|^2\} \leq \frac{\beta_n^{-1}}{n} \frac{\rho C_{\bar{u}}^2 + \sigma_{i,\bar{v}}^2}{2\sigma_{\Delta}^2 - 1} + o\left(\frac{\beta_n^{-1}}{n}\right), \quad (9)$$

Algorithm 2. Leave-out Sign-dominant Correlation Regions (LSCR) method [M. Campi and E. Weyer, TAC, 2010]

- 1 $\hat{y}_{sn+k-1}(\theta) = \Delta_n \theta^k + \sum_{i=0}^{k-1} \theta^{k-i} \bar{u}_{sn+k-l-i}$.
- 2 $\varepsilon_t(\theta) = y_t - \hat{y}_t(\theta)$, $t = 1, \dots, N$.
- 3 $f_{sn+k-1}(\theta) = \Delta_n \varepsilon_{sn+k-1}(\theta)$, $n = 0, \dots, N_\Delta - 1$, $k = 1, \dots, s$.
- 4 To choose $M > 2s$ and different random binary strings $(h_{i,1}, \dots, h_{i,N})$, $i = 0, \dots, M-1$.
To calculate $g_i^k(\theta) = \sum_{n=0}^{N_\Delta-1} h_{i,ns+k} \cdot f_{ns+k-1}(\theta)$, $i = 0, \dots, M-1$,
- 5 To choose $q \in [1; M/2s]$ and for $k = 1, \dots, s$ to build the regions:

$$\hat{\Theta}^{(k)} = \{\theta : \text{at least } q \text{ of } g_i^k(\theta) \text{ are } > 0 \text{ and at least } q \text{ are } < 0\}.$$

$$\hat{\Theta} = \bigcap_{k=1}^s \hat{\Theta}^{(k)}. \quad (10)$$

Theorem

Let condition **A1** be satisfied. Consider $k \in \{1, 2, \dots, s\}$ and assume that $\text{Prob}(g_i^k(\theta_*) = 0) = 0$. **Then**

$$\text{Prob}\{\theta_* \in \hat{\Theta}^{(k)}\} = 1 - 2q/M, \quad (11)$$

where M , q and $\hat{\Theta}^{(k)}$ is taken from 4 and 5 steps.

Theorem

$$\text{Prob}\{\theta_* \in \hat{\Theta}\} \geq 1 - 2sq/M. \quad (12)$$

Fix $k \in [1, \dots, s]$. Let H be a stochastic $M \times N_\Delta$ matrix with elements $h_{i,ns+k}$, $i = 0, 1, \dots, M-1$, $n = 0, \dots, N_\Delta - 1$, from step 4 of the algorithms in Section VI, and let $\eta = \text{col}(\eta_1, \dots, \eta_{N_\Delta})$ be a vector independent of H , consisting of mutually uncorrelated random variables symmetrically distributed around zero. Given an $i \in [0, M-1]$, let H_i be the $M \times N$ matrix, whose rows are equal to the i -th row of H . **Then**, $H\eta$ and $(H - H_i)\eta$ have the same M -dimensional distribution provided that the i -th element of $(H - H_i)\eta$ (which is 0) is repositioned as the first element of the vector.

Sketch of the Proof

Denote $\eta_n := \Delta_{n-1} \varepsilon_{(n-1)s+k-1}(\theta_*)$.

For the correlation between η_i and η_j , $i > j$,

$$E[\eta_i \eta_j] = E[\Delta_{i-1}] E[\varepsilon_{(i-1)s+k-1}(\theta_*) \Delta_{j-1} \varepsilon_{(j-1)s+k-1}(\theta_*)] = 0$$

$E[\Delta_{i-1}] = 0$ ($\eta_1, \dots, \eta_{N_\Delta}$ are mutually uncorrelated).

Take $g_{\bar{i}}^{(k)}(\theta_*)$ in the r -th position.

$$g_i^{(k)}(\theta_*) - g_{\bar{i}}^{(k)}(\theta_*) = \sum_{n=0}^{N_\Delta} (h_{i,ns+k} - h_{\bar{i},ns+k}) \eta_n < 0$$

for $r-1$ selection of $i \in [0, M-1]$.

From Proposition 1: $Prob\{\text{"}r-1\text{ entries of } (H - H_{\bar{i}})\eta \text{ are negative"}\} = Prob\{\text{"}r-1\text{ entries of } H\eta \text{ are negative"}\}$, and it does not depend on \bar{i} .

Example 1

$$y_t - 2y_{t-1} + y_{t-2} = b_* u_{t-1} + 1.6u_{t-2} + v_t, \quad t = 1, \dots, 15,$$

$y_0 = y_{-1} = u_{-1} = 0$, b_* is an unknown coefficient, v_t is an unknown external arbitrary noise.

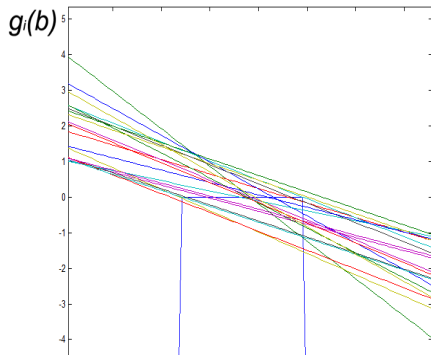
LSCR

$$b_* = 1, \quad E v_t = 0.5, \quad \sigma_v = 0.1$$

$$g_i(b) = \sum_{t=1}^{15} h_{i,t} \cdot \Delta_{t-1} \varepsilon_t(b),$$

$$h_{i,t} \in \{0, 1\}, \quad i = 1, \dots, 19.$$

The confidence interval is
[0.834; 1.090]



Example 2

$$y_t + a_*^1 y_{t-1} + y_{t-2} = b_*^1 u_{t-1} + 1,6 u_{t-2} + v_t, \quad t = 1, 2, \dots, 960,$$

$$\text{Prob}\{\theta_* \in \hat{\Theta}^{(k)}\} = 1 - 2q/M,$$

$$95\% = (1 - 2 \cdot 2 \cdot 6/480) \cdot 100\%.$$

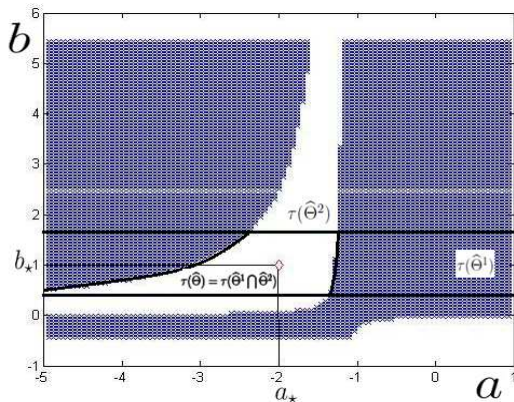


Diagram of Adaptive Control

