

Рекуррентные модификации МНК

Олег Николаевич Граничин

Санкт-Петербургский государственный университет,
математико-механический факультет

06 марта 2013

Постановка задачи

Рассмотрим задачу о построении оценки $\hat{\mathbf{x}}_N \in \mathbb{R}^r$, минимизирующей определенный ранее эмпирический функционал качества $f_N(\mathbf{x})$ при блочно-диагональной матрице весовых коэффициентов R , с блоками R_1, R_2, \dots, R_n размерности $p \times p$: $N = np$. Объединив в наборы по p наблюдения и в $(r \times p)$ -матрицы соответствующие входы модели, переобозначив их опять y_i и φ_i , выражение для функционал качества можно переписать в виде

$$f_n(\mathbf{x}) = \sum_{k=1}^n (\varphi_k^T \mathbf{x} - y_k)^T R_k (\varphi_k^T \mathbf{x} - y_k),$$

где матрицы φ_i и векторы y_i можно интерпретировать как данные наблюдения в момент времени $i = 1, 2, \dots, n$.

Вывод рекуррентного соотношения

Вектор–градиент функции $f_n(\mathbf{x})$ можно вычислить

$$\nabla f_n(\mathbf{x}) = 2 \sum_{k=1}^n \varphi_k R_k (\varphi_k^T \mathbf{x} - y_k).$$

После второго дифференцирования получаем независящую от \mathbf{x} матрицу–гессиан

$$H_n = \nabla^2 f_n(\mathbf{x}) = 2 \sum_{k=1}^n \varphi_k R_k \varphi_k^T.$$

Если оценка $\hat{\mathbf{x}}_{n-1}$ обеспечивает наименьшее значение функционалу качества $f_{n-1}(\mathbf{x})$, то $\nabla f_{n-1}(\hat{\mathbf{x}}_{n-1}) = 0$. По формуле Тейлора для $\nabla f_n(\hat{\mathbf{x}}_n)$ имеем

$$\begin{aligned} 0 = \nabla f_n(\hat{\mathbf{x}}_n) &= \nabla f_n(\hat{\mathbf{x}}_{n-1}) + H_n(\hat{\mathbf{x}}_n - \hat{\mathbf{x}}_{n-1}) = \nabla f_{n-1}(\hat{\mathbf{x}}_{n-1}) + \\ &+ 2\varphi_n R_n (\varphi_n^T \hat{\mathbf{x}}_{n-1} - y_n) + H_n(\hat{\mathbf{x}}_n - \hat{\mathbf{x}}_{n-1}). \end{aligned}$$

Обозначив $\Gamma_n = (\frac{1}{2}H_n)^{-1}$, легко выписать формулу для вычисления следующей оценки по $\hat{\mathbf{x}}_{n-1}$:

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \boldsymbol{\varphi}_n R_n (\boldsymbol{\varphi}_n^T \hat{\mathbf{x}}_{n-1} - y_n).$$

Используя матричное тождество, для Γ_n можно вывести рекуррентную формулу

$$\Gamma_n = \Gamma_{n-1} - \Gamma_{n-1} \boldsymbol{\varphi}_n (R_n^{-1} + \boldsymbol{\varphi}_n^T \Gamma_{n-1} \boldsymbol{\varphi}_n)^{-1} \boldsymbol{\varphi}_n^T \Gamma_{n-1}.$$

Полученные рекуррентные соотношения для пересчета $\hat{\mathbf{x}}_n$ и Γ_n называются *обобщенным рекуррентным методом наименьших квадратов*.

Пусть матрицы $A, D, A + B^TDB$ — невырождены.

Тогда

$$(A + B^TDB)^{-1} = A^{-1} - A^{-1}B^T(D^{-1} + BA^{-1}B^T)^{-1}BA^{-1}.$$

Проверить выполнение этого тождества можно, умножив обе части формулы на $(A + B^TDB)$.

Чтобы воспользоваться рекуррентной формой получения оценок, следует некоторым образом задать начальные значения \hat{x}_0 и Γ_0 . При их произвольном выборе, определяемые полученными рекуррентными формулами оценки, вообще говоря, не обязаны обеспечивать минимум соответствующим функционалам качества. При задании начальных значений естественно выбирать обратимую матрицу Γ_0 .

Для приложений наиболее интересен случай, когда $p = 1$, т. е. y_k и $R_k > 0$ — скалярные величины, а φ_k , $k = 1, 2, \dots$ — векторы. Эмпирический функционал качества в этом случае имеет вид

$$f_n(\mathbf{x}) = \sum_{k=1}^n R_k (\varphi_k^T \mathbf{x} - y_k)^2$$

и формулы для записи оценок обобщенного рекуррентного МНК с числовыми положительными весовыми коэффициентами:

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \varphi_n R_n (\varphi_n^T \hat{\mathbf{x}}_{n-1} - y_n),$$

$$\Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1} \varphi_n \varphi_n^T \Gamma_{n-1}}{R_n^{-1} + \varphi_n^T \Gamma_{n-1} \varphi_n}.$$

Полагая $R_k = 1$, $k = 1, 2, \dots$, получаем формулы обыкновенного рекуррентного МНК

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \boldsymbol{\varphi}_n (\boldsymbol{\varphi}_n^T \hat{\mathbf{x}}_{n-1} - y_n), \quad \hat{\mathbf{x}}_0 = 0,$$

$$\Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1} \boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^T \Gamma_{n-1}}{1 + \boldsymbol{\varphi}_n^T \Gamma_{n-1} \boldsymbol{\varphi}_n}, \quad \Gamma_0 = \gamma_0^{-1} \mathbf{I},$$

где $\gamma_0 > 0$ — малый параметр регуляризации. Этот алгоритм — частный вид *фильтра Калмана–Бьюси*.

В случае независимых наблюдений и гауссовых помех он обладает оптимальными свойствами.

Во многих задачах эффективно используют *модифицированный МНК* с “забывающим” множителем, получающийся при функционале качества с $R_k = \gamma^{N-k}$, $k = 1, 2, \dots, N$,

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \boldsymbol{\varphi}_n (\boldsymbol{\varphi}_n^T \hat{\mathbf{x}}_{n-1} - y_n), \quad \hat{\mathbf{x}}_0 = 0,$$

$$\Gamma_n = \gamma^{-1} \left(\Gamma_{n-1} - \frac{\Gamma_{n-1} \boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^T \Gamma_{n-1}}{\gamma + \boldsymbol{\varphi}_n^T \Gamma_{n-1} \boldsymbol{\varphi}_n} \right), \quad \Gamma_0 = \gamma^{-1} \mathbf{I},$$

где $\gamma \in (0, 1)$ — забывающий множитель.

В той ситуации, когда нельзя предполагать независимость наблюдений, при синтезе адаптивного управления стохастическим линейным объектом используется рекуррентный МНК, соответствующий выбору весовых коэффициентов

$$R_n = \frac{1}{\ln(\gamma^{-1} + \sum_{i=1}^n \|\varphi_i\|^2)}.$$

Условие “постоянного возбуждения”

Матрицы Γ_n монотонно убывают если выполнено условие *постоянного возбуждения* для последовательности $\{\varphi_n\}$: существует целое число $\bar{N} > 0$ и постоянная $\delta > 0$ такие, что

$$\sum_{k=n}^{n+\bar{N}} E\{\varphi_k \varphi_k^T\} \geq \delta I, \quad n = 1, 2, \dots$$

Рекуррентная процедура МНК, являясь оптимальной при соответствующем выборе начальных статистик, становится практически малопригодной, если приходится оценивать вектор параметров высокой размерности: основной объем вычислений связан с пересчетом матриц Γ_n . Естественно попытаться её упростить, даже если придется поступиться оптимальными свойствами. Впрочем, последнее обстоятельство не является определяющим, так как нужные начальные данные обычно неизвестны, а произвольный выбор начальных данных делает процедуру только предельно оптимальной.

Следующие упрощения рекуррентной процедуры МНК в типичных случаях не сказываются на предельных свойствах оценок.

Если $\Gamma_n \rightarrow 0$ при $n \rightarrow \infty$, то в том случае, когда φ_n равномерно ограничены, при достаточно больших n справедливо:

$(R_n^{-1} + \varphi_n^T \Gamma_{n-1} \varphi_n)^{-1} \approx R_n$. Следовательно, рекуррентный алгоритм для пересчета оценок можно записать в виде

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \varphi_n R_n (\varphi_n^T \hat{\mathbf{x}}_{n-1} - y_n),$$

$$\Gamma_n = \Gamma_{n-1} - \Gamma_{n-1} \varphi_n R_n \varphi_n^T \Gamma_{n-1}.$$

Другим существенным упрощением в рекуррентной процедуре МНК является использование в алгоритме вместо матриц Γ_n числовых коэффициентов.

При предположении, что входы $\{\varphi_n\}$ имеют статистическую природу и известны их средние значения, в условиях почти произвольных помех в наблюдениях обоснована состоятельность оценок *рандомизированного рекуррентного МНК*, имеющего при $\gamma_0 > 0$ вид

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_{n-1} - \Gamma_n \Delta_n (\varphi_n^T \hat{\mathbf{x}}_{n-1} - y_n), \quad \hat{\mathbf{x}}_0 = 0,$$

$$\Gamma_n = \Gamma_{n-1} - \frac{\Gamma_{n-1} \Delta_n \Delta_n^T \Gamma_{n-1}}{1 + \Delta_n^T \Gamma_{n-1} \Delta_n}, \quad \Gamma_0 = \gamma_0^{-1} \mathbf{I},$$

где $\Delta_n = \varphi_n - E\{\varphi_n\}$, $n = 1, 2, \dots$ — центрированные входы.