3.36pt

Статистические методы регуляризации

Факультет Космических Исследований, МГУ

Москва

Постановка задачи в матричном виде

Описание измерений

$$y = F(x) + \varepsilon,$$

Матрица весовых функций (weighting function matrix)
Линеаризация приводит к записи:

$$y - F(x_0) = \frac{\partial F(x)}{\partial x}(x - x_0) + \epsilon = K(x - x_0) + \epsilon,$$

где матирца K размера $m \times n$.
Ранг матрицы p, существует нулевое пространство, если $p < n$ - нeединвтенность решения, Задача переопределена, если $m > p$,

Эксперимент с измеряемой ошибкой

$$\bar{y} = \int yP(y)dy,$$

$$\sigma^2 \int (y - \bar{y})^2 P(y)dy.$$

В качестве распределения вероянтности берется Гаусово нормальное распределение

$$P(y) = N(y - \bar{y}, \sigma) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\{-\frac{(y - \bar{y})^2}{2\sigma^2}\}$$

Ковариационная матрица

$$S_{ij} = \mathcal{E}\{(y_i - \bar{y}_i)(y_j - \bar{y}_j)\},$$

Гаусово распределения для вектора

$$P(y) = \frac{1}{(2\pi)^{n/2}|S_y|^{1/2}} \exp\{-\frac{1}{2}(y - \bar{y})^T S_y^{-1}(y - \bar{y})\}$$

Байсовский подход к решению обратных задач

Пусть известны вероятности измерений $P(x)$, $P(y)$, совмещенная вероятность, также заданы условные вероятности $P(y|x)$, $P(x|y)$. Тогда справедливы равенства
$P(x) = \int P(x,y)dy$ Формула условной вероятности

$$P(y|x) = P(x,y)/P(x)$$

так что для постериорной информации можно получить формулу:

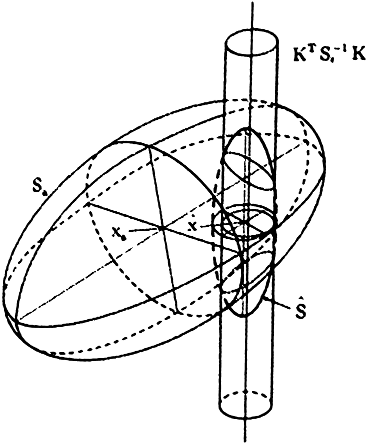
$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

Основные положения

1. До измерения у нас есть a priori информация;
2. Процесс измерений представлен прямой задачей, которая переводит ппространство состояний в пространство измерений;
3. Процедура обращения основана на Байесовом формализме;

Иллюстрация в верноятностном пространстве

Отображение области a priori данных, измерения в пространство решений, для трехмерного просранства решений и двумерного пространства измерений.



Основные представления решения

На основании представления

$$-2\ln P(x|y) = (y-Kx)^T S_\epsilon^{-1}(y-Kx) + (x-x_a)^T S_a^{-1}(x-x_a),$$

используя предположении о Гауссовом распределении решения

$$-2\ln P(x|y) = (x-\bar{x})^T \bar{S}^{-1}(x-\bar{x}).$$

получаем следующие два возможные представления для решения:

$$\hat{x} = x_a + (K^T S_\epsilon^{-1} K + S_a^{-1})^{-1} K^T S_\epsilon^{-1}(y - Kx_a),$$

$$\hat{x} = x_a + S_a K^T (K S_a K^T + S_\epsilon)^{-1}(y - Kx_a),$$

Степени свободы решения

Рассмотрим простейший пример:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix},$$

что эквивалентно измерению двух ортогональных величин

$$z_1 = y_1 + y_2 = 2(x_1 + x_2) + \epsilon_1 + \epsilon_2,$$

$$z_2 = y_1 - y_2 = 0.02(x_1 - x_2) + \epsilon_1 - \epsilon_2.$$

Нормировка шумов

Для задачи $y = K(x - x_a) + \epsilon$ проведем замену переменных:

$$\tilde{x} = S_a^{-\frac{1}{2}}(x - x_a),\ \tilde{y} = S_\epsilon^{-\frac{1}{2}} y$$

$$\tilde{y} = S_\epsilon^{-\frac{1}{2}} K S_a^{\frac{1}{2}} \tilde{x} + = S_\epsilon^{-\frac{1}{2}} \epsilon = \tilde{K}\tilde{x} + \tilde{\epsilon}.$$

После сингулярного разложения ядра $\tilde{K} = U\Lambda V^T$:

$$y' = U^T \tilde{y} = U^T \tilde{K} \tilde{x} + U^T \tilde{\epsilon} = \Lambda V^T \tilde{x} + \epsilon' = \Lambda x' + \epsilon'.$$

Так что элементы y' которые определены более чем шумы - это те для которых диагональные элементы превосходят единицу $\lambda \succeq 1$.

Степени свободы решения 1

Определим сколько степеней свободы относятся к решению, а сколько к шумам

$$\hat{x} = \frac{\sigma_a^2 x + \sigma_\epsilon^2 x_a}{\sigma_a^2 + \sigma_\epsilon^2},$$

$$\xi^2 = (x - x_a)^T S_a^T (x - x_a) + (y - Kx)^T S_\epsilon^{-1} (y - Kx),$$

Таким образом, степени свободы решения и шумов определяются величинами, соответственно:

$$d_s = \mathcal{E}\{(\bar{x} - x_a)^T S_a^{-1} (\bar{x} - x_a)\},$$

$$d_n = \mathcal{E}\{\bar{\epsilon} S_\epsilon^{-1} \bar{\epsilon}\}$$

Степени свободы решения 2

Имеет место выражение $d_s + d_n = tr(I_m)$. Замечаем, что одна из форма для $d_s = tr(GK)$. (NB матриц $A = GK$ средних ядер)

$$\hat{x} - x_a = Gy = G[K(x - x_a) + \epsilon] = A(x - x_a) + G\epsilon.$$

В результате можно получить простые формулы для расчета степеней сигнала и шумов:

$$d_s = tr(\Lambda^2(\Lambda^2 + I_m)^{-1}) = \sum \lambda_i^2/(1 + \lambda_i^2),$$

$$d_n = tr((\Lambda^2 + I_m)^{-1}) = \sum 1/(1 + \lambda_i^2),$$

Анализ ошибок

Прямая функция

$$y = f(x, b) + \epsilon,$$

обратная процедура восстановления

$$\bar{x} = R(y, b, x_a, c).$$

Пусть модель будет построена с некоторыми допущениями (b'-опущеные параметры исходной функции)

$$\triangle f = f(x, b, b') - F(x, b),$$

Линеаризуем исходную задачу в окрестности априорных значениях x_a:

$$\bar{x} = R(F(x_a, \tilde{b}) + K_x(x - x_a) + K_b(b - b') + \triangle f(x, b, b') + \epsilon, \tilde{b}, x_a, c)$$

Средние ядра

Продифференцируем обратный оператор:

$$\bar{x} = R(F(x_a, \tilde{b}), \tilde{b}, x_a, c) + G_y[K_x(x - x_a) + K_b(b - b') + \triangle f(x, b, b') + \epsilon],$$

Таким образом можно определить средние ядра как

$$A = G_y K_x = \frac{\partial \tilde{x}}{\partial \tilde{x}},$$

другими словами - чувствительность получаемого решения к точным данным.

$$K_b(b - b') + \triangle f(x, b, b') + \epsilon,$$

- полная ошибка восстановления к прямой задаче.

Вертикальное разрешение 1

1. Разрешение позволяет понять, как хорошо мы можем видеть атмосферные особенности. Отклик на возмущение отдельно взятой величины, а также диапазон высот, на которые оказывает влияние отдельно взятая величина.
2. Разрешение надо отличать от размерности сетки.
3. Измерение разрешения имеет смысл, когда есть оценка ошибки точности.

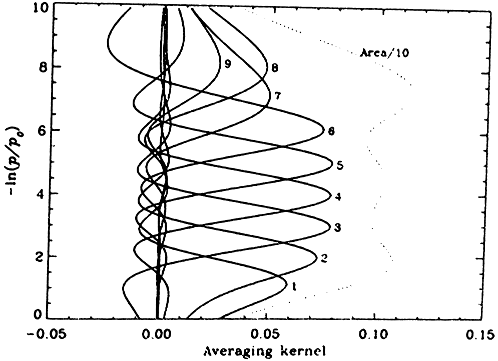
Степень свободы сигнала d_s - это tr - матрицы усредненных ядер. То есть диагональ матрица A можно рассматривать как меру степеней свободы на каждом уровне.

Ширина усредненных ядер

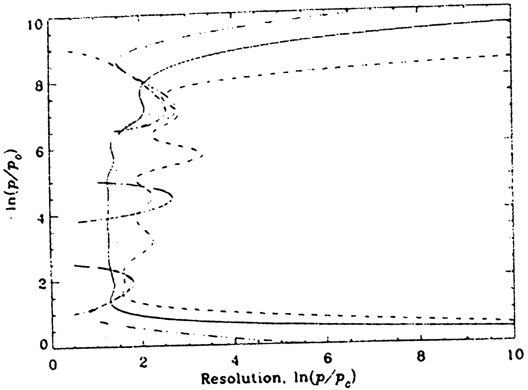
Существует несколько способов определить ширину пика функции. Можно дать определение как ширина на половине высоты. Также необходимо дать алгебраическую оценку.

1. Определим среднее значение $\bar{z} = \int z' A(z, z') dz' / \int A(z, z') dz'$, тогда
$$\varpi(z) = \left(\frac{\int A(z, z')(z' - \bar{z})^2 dz'}{\int A(z, z') dz'} \right)^(1/2).$$
2. По Бэкусу и Гильберту определяется "spread"
$$s(z) = 12 \int (z - z')^2 A^2(z, z') dz' / (\int A(z, z') dz')^2.$$
Коэффициент 12 выбран так, что щелевая функция имеет "spread"равный его полной ширине.
3. Аналгоичнй подход "resolving length"разрешающей длины
$$c(z) = \int z' A^2(z, z') dz' / \int A^2(z, z') dz'.$$

Пример усредненных ядер



Вертикальное разрешение



Линейные методы

Максимальное a posteriori решение -решение с максимальной вероятностью $P(x|y)$ можно найти как случайную величину

$$\hat{x} = \int xP(x|y)dx.$$

Для линейной проблемы с Гауссовой плотностью можно получено по методу Байсса. С матрицей восстановления

$$G = (K^T S_\epsilon^{-1} K + S_a^{-1})^{-1} K^T S_\epsilon^{-1},$$

и матрицей ошибок

$$\bar{S} = (K^T S_\epsilon^{-1} K + S_a^{-1})^{-1}.$$

Методы минимальной ошибки

Ищем метод, который минимизирует ожидаемое значение получаемой ошибки.

$$\bar{x} = x_0 + Gy.$$

Для этого рассмотрим решение по строкам

$$\mathcal{E}\{(\hat{x}_i - x_i)^2\} = \mathcal{E}\{(x_{0i} - x_i + g_i^T y)^2\}$$

Различные виды нелинейности

1. ЛИНЕЙНЫЕ когда прямая задача может быть представлена в форме $y = Kx$ и априорная информация имеет Гаусово распределение
2. ПОЧТИ ЛИНЕЙНЫЕ разновидность нелинейных задач, для которых линеаризация в окрестности a priori данных. Задачи которые линейные в пределах точности измерений или в ожидаемой точности решений
3. УМЕРЕННО НЕЛИНЕЙНЫЕ у которых линеаризации достаточно для анализа ошибок, но не для нахождения решения.
4. СУЩЕСТВЕННО НЕЛИНЕЙНЫЕ задачи нелинейные даже в пределах допутсимых ошибок.

Определение степени нелинейности

$$\delta\hat{x} = G[F(\hat{x}) - F(x_a) - K(x - x_a)].$$

Опитмальный подход для нелинейных моделей

Байесовское решение для линейной задачи может быть заменено напрямую для прямой задачи с произвольной функцией

$$-2\ln P(x|y) = [y - F(x)]^T S_\epsilon^{-1}[y - F(x)] + [x - x_a]^T S_a^{-1}[x - x_a],$$

для нахождения наиболее вероятного состояния, проидифференцируем

$$\nabla -2\ln P(x|y) = -[\nabla_x F(x)]^T S_\epsilon^{-1}[y - F(x)] + S_a^{-1}[x - x_a] = 0,$$

Методы Ньютона и Гаусса-Ньютона

Метод Ньютона для скалярного случая можно записать:

$$x_{i+1} = x_i - [\nabla_x g(x_i)]^{-1} g(x_i),$$

$$\nabla_x g = S_a^{-1} + K^T S_\epsilon^{-1} K - [\nabla_x K^T] S_\epsilon^{-1} [y - F(x)],$$

где g - производная исходной степенной функции, а $\nabla_x g$ - вторая производная известная также как гессиан.

$$x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i)^{-1} [K_i^T S_\epsilon^{-1} (y - F(x)) - S_a^{-1}(x_i - x_a)].$$

Распространенная ошибка

Итерационный процесс в общем виде:

$$x_{i+1} = x_a + G_i[y - F(x_i) + K_i(x_i - x_a)],$$

Может показаться, что использование a priori значения в качестве уточнения для итерационного процесса:

$$x_{i+1} = x_a + G_i[y - F(x_i)],$$

В предельном случае:

$$\bar{x} = \bar{x} + \bar{G}[y - F(\bar{x})].$$

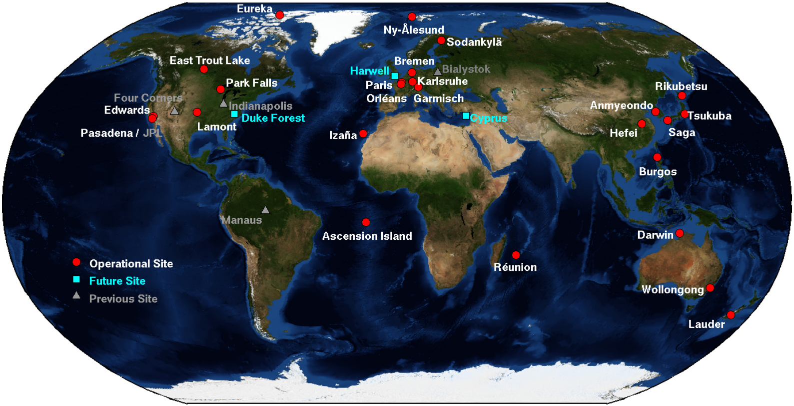
Метод Левенберга-Марквардта

Метод Ньютона является точным для квадратичной функции. Если первое приближение достаточно далеко от точного, то квадратичная апроксимация медленно приводит к решению или вообще уводить от него.

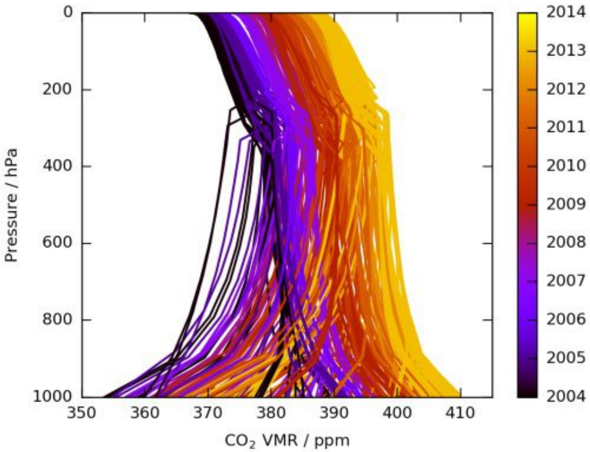
$$x_{i+1} = x_i + (KK^T + \gamma_i I)^{-1} K_T [y - F(x_i)],$$

где парамтр γ_i выбирается так, чтобы минимизировать весовую функцию.
При $\gamma_i \to 0$ итерации близки к Гаусс-Ньютону, а для $\gamma_i \to \infty$ итерационный процесс близок к наискорейшому спуску. Уменьшения невязки при достаточно больших параметров, но возникнет эффект овражности. Меняем единичную матрицу на диагональ матрицы гессе, тогда будет учитываться кривизна функционала.

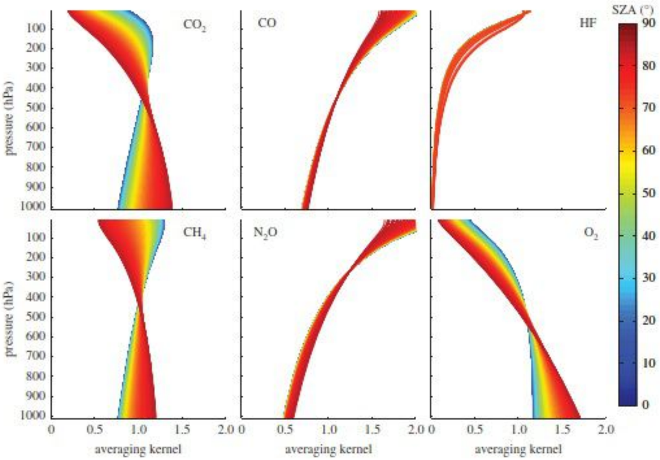
Измирительная сеть TCCON



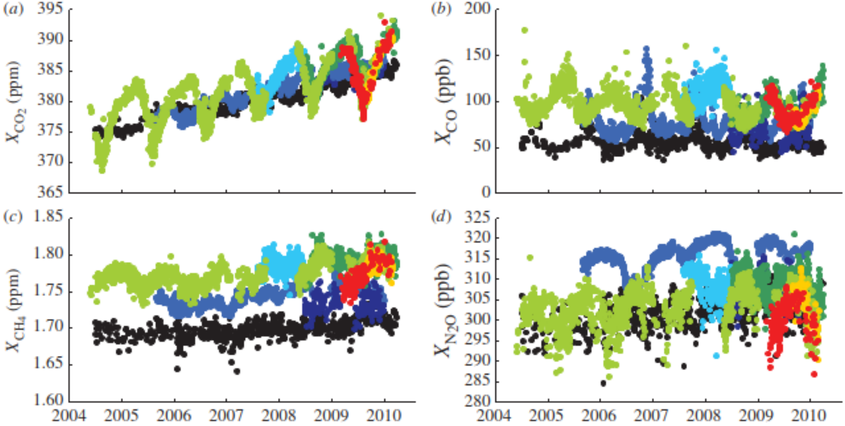
A priori профили CO2

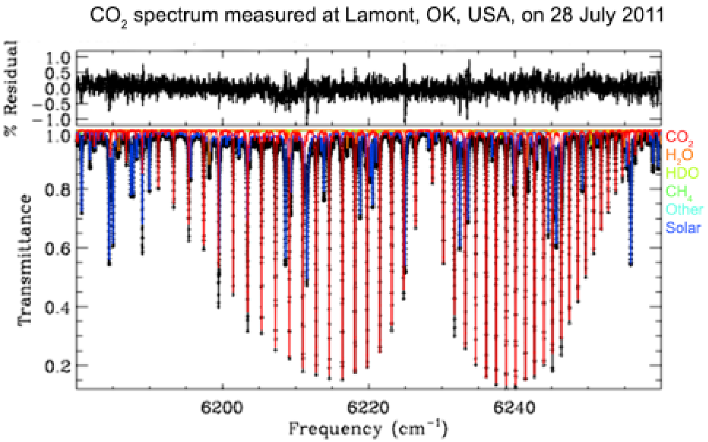


A priori профили различных составляющих

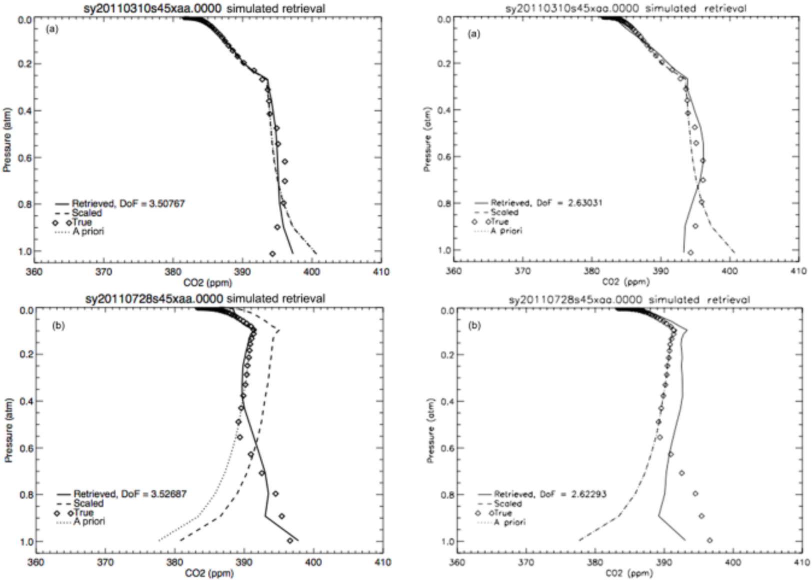


Тренды различных газовых составляющих




CO2 spectrum measured at Lamont, OK, USA, on 28 July 2011
% Residual
1.0
0.5
0.0
−0.5
−1.0
Transmittance
1.0
0.8
0.6
0.4
0.2
6200
6220
6240
Frequency (cm⁻¹)
CO2
H2O
HDO
CH4
Other
Solar

Примеры восстановления профилей co2 1



Примеры восстановления профилей co2 2