ДЗ 5 по Методам Оптимизации. Векторное дифференцирование

Соколов Игорь, группа 573

10 октября 2017 г.

1

Найти $\nabla f(x)$, если $f(x) = \|Ax\| - \|x^T A\|$

Решение:

Перепишем в скалярном виде:

$$f(x) = \sqrt{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2} - \sqrt{\sum_{j=1}^{n}\left(\sum_{i=1}^{n} x_i a_{ij}\right)^2}$$

$$\frac{\partial f}{\partial x_k} = \frac{\sum\limits_{i=1}^{n} 2a_{ik}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)}{2\sqrt{\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)^2}} - \frac{\sum\limits_{j=1}^{n} 2a_{kj}\left(\sum\limits_{i=1}^{n} a_{ij}x_j\right)}{2\sqrt{\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)^2}} =$$

$$= \frac{\sum\limits_{i=1}^{n} a_{ik}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)}{\sqrt{\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)^2}} - \frac{\sum\limits_{j=1}^{n} a_{kj}\left(\sum\limits_{i=1}^{n} a_{ij}x_j\right)}{\sqrt{\sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} a_{ij}x_j\right)^2}} =$$

$$= \frac{a_k^T\left(Ax\right)}{\|Ax\|} - \frac{a_k(x^T A)^T}{\|x^T A\|} = \frac{a_k^T\left(Ax\right)}{\|Ax\|} - \frac{a_k(A^T x)}{\|x^T A\|} \quad (1)$$

$$\nabla f(x) = \frac{A^T\left(Ax\right)}{\|Ax\|} - \frac{A(A^T x)}{\|x^T A\|}$$

Ответ: $\nabla f(x) = \dfrac{A^T\left(Ax\right)}{\|Ax\|} - \dfrac{A(A^T x)}{\|x^T A\|}$

2

Найти $\nabla f(x), f''(x)$, если $f(x) = \dfrac{-1}{1+x^T x}$

Решение:

Перепишем $f(x)$ в скаляром виде:

$$f(x) = \frac{-1}{1+\sum\limits_{i=1}^{n} x_i^2}$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\frac{-1}{1+\sum\limits_{i=1}^{n} x_i^2}\right) = \frac{2x_k}{\left(1+\sum\limits_{i=1}^{n} x_i^2\right)^2}$$

$$\Rightarrow \nabla f(x) = \frac{2x}{\left(1+x^T x\right)^2}$$

Пусть $g_k = \dfrac{\partial f(x)}{\partial x_k} \to H_{k,p} = \dfrac{\partial g_k}{\partial x_p}$

Тогда

$$H_{k,p} = \frac{\partial^2 f(x)}{\partial x_p \partial x_k} = \frac{\partial}{\partial x_p}\left(\frac{2x_k}{\left(1+\sum\limits_{i=1}^{n} x_i^2\right)^2}\right) =$$

$$= \frac{2\delta_{k,p}\left(1+\sum\limits_{i=1}^{n} x_i^2\right)^2 - 8x_k x_p\left(1+\sum\limits_{i=1}^{n} x_i^2\right)}{\left(1+\sum\limits_{i=1}^{n} x_i^2\right)^4} = \frac{2}{\left(1+\sum\limits_{i=1}^{n} x_i^2\right)^2}\left(\delta_{k,p} - \frac{4x_p x_k}{1+\sum\limits_{i=1}^{n} x_i^2}\right) \quad (1)$$

$$\Rightarrow f''(x) = \frac{2}{\left(1+x^T x\right)^2}\left(\mathbb{E} - \frac{4xx^T}{1+x^T x}\right)$$

Ответ:

$\nabla f(x) = \dfrac{2x}{\left(1+x^T x\right)^2}$

$f''(x) = \dfrac{2}{\left(1+x^T x\right)^2}\left(\mathbb{E} - \dfrac{4xx^T}{1+x^T x}\right)$

3

Найти $f'(X)$, если $f(X) = \det X$

Примечание: здесь под $f'(X)$ подразумевается оценка фунции $f(X)$ первого порядка в смысле разложения в ряд Тейлора:

$$f(X + \Delta X) \approx f(X) + \mathbf{tr}(f'(X)^T \Delta X)$$

Решение:

Пусть $g(X) = \log\det X$

Найдем $g(X + \Delta X) = \log\det X$, где ΔX - некоторая малая матрица по норме Фробениуса ( ).

Тогда $f(X + \Delta X) = e^{g(X+\Delta X)} = \det(X + \Delta X)$

Вспомним, что производная имеет смысл линейной аппроксимация функции в окрестности точки.

Заметим, что:

$$\begin{aligned}\log\det\left[X + \Delta X\right] &= \log\det\left[X^{1/2}\left(I + X^{-1/2}\Delta X X^{-1/2}\right)X^{1/2}\right] = \\ &= \log\det\left[X^{1/2}\right]\det\left[I + X^{-1/2}\Delta X X^{-1/2}\right]\det\left[X^{1/2}\right] = \\ &= \log\det\left[X\right]\det\left[I + X^{-1/2}\Delta X X^{-1/2}\right] = \\ &= \log\det\left[X\right] + \log\det\left[I + X^{-1/2}\Delta X X^{-1/2}\right] \quad (1)\end{aligned}$$

Известно, что *определитель матрицы равен произведению её собственных значений*

$$\log\det\left[X + \Delta X\right] = \log\det X + \sum_{i=1}^{n}\log(1 + \lambda_i)$$

Где λ_i - собственные числа матрицы $X^{-1/2}\Delta X X^{-1/2}$. Далее используем факт "малости"матрицы ΔX (в смысле нормы этой матрицы), следовательно для приближения первого порядка справедливо: $\log(1 + \lambda_i) \approx \lambda_i$ т.к. λ_i так же должны быть малыми.

$$\log\det\left[X + \Delta X\right] \approx \log\det X + \sum_{i=1}^{n}\lambda_i$$

Также пользуемся свойством следа матрицы: *след матрицы равен сумме её собственных значений.*

$$\begin{aligned}\log\det\left[X + \Delta X\right] &\approx \log\det X + \mathbf{tr}\left[X^{-1/2}\Delta X X^{-1/2}\right] = \\ &= \log\det X + \mathbf{tr}\left[X^{-1/2}X^{-1/2}\Delta X\right] = \log\det X + \mathbf{tr}\left[X^{-1}\Delta X\right] \quad (2)\end{aligned}$$

Так как по условию матрица X - квадратная, симметричная, то $(X^{-1})^T = \left(X^T\right)^{-1} = X^{-1}$.

Заметим, что след произведения матриц есть обобщения скалярного произведения на пространство матриц.

След матрицы также обладает следующий свойством:

$$\mathbf{tr}(A^TB) = \mathbf{tr}(AB^T) = \mathbf{tr}(B^TA) = \mathbf{tr}(BA^T)$$

.

В силу всех вышеприведенных фактов, справедливо:

$$g(X + \Delta X) \approx g(X) + \langle g'(X), \Delta X\rangle \tag{3}$$

$$\Rightarrow g'(X) = X^{-1} \tag{4}$$

Из (2) $\Rightarrow$

$$e^{\log\det[X+\Delta X]} = e^{\log\det X + \mathbf{tr}\left[X^{-1}\Delta X\right]}$$
$$\det\left[X + \Delta X\right] = e^{\log\det X} e^{\mathbf{tr}\left[X^{-1}\Delta X\right]}$$

В силу малости матрицы ΔX справедливо $e^{\mathbf{tr}\left[X^{-1}\Delta X\right]} = 1 + \mathbf{tr}\left[X^{-1}\Delta X\right]$

$$\Rightarrow \det\left[X + \Delta X\right] = \det X\left(1 + \mathbf{tr}\left[X^{-1}\Delta X\right]\right)$$
$$\det\left[X + \Delta X\right] = \det X + \det X\,\mathbf{tr}\left[X^{-1}\Delta X\right] \tag{5}$$

Заметим, что $detX$ - скаляр и в силу свойства линейности следа матрицы

$$\mathbf{tr}\,(\alpha A) = \alpha\,\mathbf{tr}\,(A)$$

Получаем

$$\det\left[X + \Delta X\right] = \det X + \mathbf{tr}\left[(\det X)X^{-1}\Delta X\right]$$

Пользуясь аналогичными примеру (3) рассуждениями, получаем:

$$\det\left[X + \Delta X\right] = \det X + \mathbf{tr}\langle(\det X)X^{-1}, \Delta X\rangle$$

Значит, $f'(X) = det(X)X^{-1}$

Ответ: $f'(X) = det(X)X^{-1}$

4

Найти $f''(X)$, если $f(X) = \log\det X$

Примечание: здесь под $f''(X)$ подразумевается оценка фунции $f(X)$ второго порядка в смысле разложения в ряд Тейлора:

$$f(X + \Delta X) \approx f(X) + \mathbf{tr}(f'(X)^T\Delta X) + \frac{1}{2}\mathbf{tr}(\Delta X^T f''(X)\Delta X)$$

Решение:

Задача решается аналогично задаче 3.
Воспользуемся примечанием к задаче.

$$\begin{aligned}\log\det\left[X+\Delta X\right] &= \log\det\left[X^{1/2}\left(I+X^{-1/2}\Delta X X^{-1/2}\right)X^{1/2}\right] = \\ &= \log\det\left[X^{1/2}\right]\det\left[I+X^{-1/2}\Delta X X^{-1/2}\right]\det\left[X^{1/2}\right] = \\ &= \log\det\left[X\right]\det\left[I+X^{-1/2}\Delta X X^{-1/2}\right] = \\ &= \log\det\left[X\right]+\log\det\left[I+X^{-1/2}\Delta X X^{-1/2}\right] \quad (1)\end{aligned}$$

Известно, что *определитель матрицы равен произведению её собственных значений*

$$\log\det\left[X+\Delta X\right] = \log\det X + \sum_{i=1}^{n}\log(1+\lambda_i)$$

Где λ_i - собственные числа матрицы $X^{-1/2}\Delta X X^{-1/2}$. Далее используем факт "малости"матрицы ΔX (в смысле нормы этой матрицы), следовательно для приближения второго порядка справедливо: $\log(1+\lambda_i) \approx \lambda_i - \dfrac{\lambda_i^2}{2}$ т.к. λ_i так же должны быть малыми.

$$\log\det\left[X+\Delta X\right] \approx \log\det X + \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\lambda_i^2$$

Докажем, что $\mathbf{tr}(XX) = \sum\limits_{i=1}^{n}\lambda_i^2$, где λ_i - собственные значения матрицы X, а $XX = X^2$.

Доказательство. Пусть S - матрица перехода из исходного базиса матрицы X к базису, в котором матрица X диагональна и на её диагонали стоят собственные значения (столбцы S есть собственные вектора матрицы X записанные в старом базисе)

Тогда $X_{eigen} = S^{-1}XS$

Воспользуемся свойством циклической перестановки:

$\mathbf{tr}(ABC) = \mathbf{tr}(BCA) = \mathbf{tr}(CAB)$

$\mathbf{tr}(X_{eigen}) = \sum\limits_{i=1}^{n}\lambda_i = \mathbf{tr}(S^{-1}XS) = \mathbf{tr}(SS^{-1}X) = \mathbf{tr}(X)$ (доказали ранее сформулированное свойство)

Воспользуемся еще одним свойством следа матрицы:

$\mathbf{tr}(X^TY) = \sum\limits_{i,j}X_{i,j}Y_{i,j}$

В нашем случае:

$\mathbf{tr}(X^TX) = \mathbf{tr}(XX) = \sum\limits_{i,j}X_{i,j}X_{i,j}$

$\Rightarrow \mathbf{tr}(X_{eigen}X_{eigen}) = \sum\limits_{i=1}^{n}\lambda_i^2$

Но $\mathbf{tr}(X_{eigen}X_{eigen}) = \mathbf{tr}(S^{-1}XSS^{-1}XS) = \mathbf{tr}(S^{-1}XXS) = \mathbf{tr}(SS^{-1}XX) = \mathbf{tr}(XX)$

$\Rightarrow \mathbf{tr}(XX) = \sum\limits_{i=1}^{n}\lambda_i^2$ $\square$

$$\log\det\left[X+\Delta X\right] \approx \log\det X + \mathbf{tr}\left[X^{-1/2}\Delta X X^{-1/2}\right] - \frac{1}{2}\mathbf{tr}\left[X^{-1/2}\Delta X X^{-1/2} X^{-1/2}\Delta X X^{-1/2}\right] =$$

$$= \log\det X + \mathbf{tr}\left[X^{-1/2}X^{-1/2}\Delta X\right] - \frac{1}{2}\mathbf{tr}\left[X^{-1}\Delta X X^{-1}\Delta X\right] =$$

$$= \log\det X + \mathbf{tr}\left[\left(X^{-1}\right)^T \Delta X\right] + \frac{1}{2}\mathbf{tr}\left[\Delta X^T\left(-X^{-1}X^{-1}\right)\Delta X\right] \quad (2)$$

В примере (2) воспользовались тем что X - квадратная и симметричная.
$\Rightarrow f''(X) = -\left(X^{-1}\right)^2$

Ответ: $f''(X) = -\left(X^{-1}\right)^2$

5

Найти градиент и гессиан функции $f:\mathbb{R}^n \to \mathbb{R}$, $f(x) = \log\sum\limits_{i=1}^{m}\exp(a_i^T x + b_i)$, $\quad a_1,\dots,a_m \in \mathbb{R}^n; \quad b_1,\dots,b_m \in \mathbb{R}$

Решение:

$$\frac{\partial f}{x_k} = \frac{\sum\limits_{i=1}^{m} a_{ik}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)}{\sum\limits_{i=1}^{m}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)} \quad (1)$$

$$\nabla f = \frac{\sum\limits_{i=1}^{m} a_i\exp\left(a_i^T x + b_i\right)}{\sum\limits_{i=1}^{m}\exp\left(a_i^T x + b_i\right)} \quad (2)$$

Пусть $g_k = \dfrac{\partial f}{\partial x_k}$
$H_{k,p} = \dfrac{\partial g}{\partial x_p}$
Тогда

$$H_{k,p} = \frac{\partial g}{\partial x_p} = \frac{\partial^2 f}{\partial x_p \partial x_k} =$$

$$= \frac{\sum\limits_{i=1}^{m} a_{ik}a_{ip}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)}{\sum\limits_{i=1}^{m}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)} - \frac{\left[\sum\limits_{i=1}^{m} a_{ik}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)\right]\left[\sum\limits_{i=1}^{m} a_{ik}\exp\left(\sum\limits_{j=1}^{n} a_{ip}x_j + b_i\right)\right]}{\left[\sum\limits_{i=1}^{m}\exp\left(\sum\limits_{j=1}^{n} a_{ij}x_j + b_i\right)\right]^2} \quad (3)$$

$$f''(x) = \frac{\sum\limits_{i=1}^{m} a_i a_i^T \exp\left(a_i^T x + b_i\right)}{\sum\limits_{i=1}^{m} \exp\left(a_i^T x + b_i\right)} - \frac{\left[\sum\limits_{i=1}^{m} a_i \exp\left(a_i^T x + b_i\right)\right]\left[\sum\limits_{i=1}^{m} a_i^T \exp\left(a_i^T x + b_i\right)\right]}{\left[\sum\limits_{i=1}^{m} \exp\left(a_i^T x + b_i\right)\right]^2}$$

Ответ:

$$\nabla f = \frac{\sum\limits_{i=1}^{m} a_i \exp\left(a_i^T x + b_i\right)}{\sum\limits_{i=1}^{m} \exp\left(a_i^T x + b_i\right)}$$

$$f''(x) = \frac{\sum\limits_{i=1}^{m} a_i a_i^T \exp\left(a_i^T x + b_i\right)}{\sum\limits_{i=1}^{m} \exp\left(a_i^T x + b_i\right)} - \frac{\left[\sum\limits_{i=1}^{m} a_i \exp\left(a_i^T x + b_i\right)\right]\left[\sum\limits_{i=1}^{m} a_i^T \exp\left(a_i^T x + b_i\right)\right]}{\left[\sum\limits_{i=1}^{m} \exp\left(a_i^T x + b_i\right)\right]^2}$$