Basic concepts in Linear Algebra and Optimization

Yinbin Ma

GEOPHYS 211

Outline

- Basic Concepts on Linear Algbra
  - vector space
  - norm
  - linear mapping, range, null space
  - matrix multiplication
- Iterative Methods for Linear Optimization
  - normal equation
  - steepest descent
  - conjugate gradient
- Unconstrainted Nonlinear Optimization
  - Optimality condition
  - Methods based on a local quadratic model
  - Line search methods

Outline

- Basic Concepts on Linear Algbra
  - vector space
  - norm
  - linear mapping, range, null space
  - matrix multiplication
- Iterative Methods for Linear Optimization
  - normal equation
  - steepest descent
  - conjugate gradient
- Unconstrainted Nonlinear Optimization
  - Optimality condition
  - Methods based on a local quadratic model
  - Line search methods

Outline

- Basic Concepts on Linear Algbra
  - vector space
  - norm
  - linear mapping, range, null space
  - matrix multiplication
- Iterative Methods for Linear Optimization
  - normal equation
  - steepest descent
  - conjugate gradient
- Unconstrainted Nonlinear Optimization
  - Optimality condition
  - Methods based on a local quadratic model
  - Line search methods

Basic concepts - vector space

A vector space is any set V for which two operations are defined:
1) Vector addition: any vector x_1 and x_2 in set V can be added to another vector $x = x_1 + x_2$ and x is also in set V.
2) Scalar Multiplication: Any vector x in V can be multiplied ("scaled") by a real number $c \in \mathbf{R}$ to produce a second vector cx which is also in V.

In this class, we only discuss the case where $V \subset \mathbb{R}^n$, meaning each vector x is the space is a n-dimensional column vector.

Basic concepts - norm

The “model space” and “data space” we mentioned in class are normed vector spaces. A norm is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ that map a vector to a real number. A norm must satisfy the following:
1) $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$
2) $\|x + y\| \leq \|x\| + \|y\|$
3) $\|\alpha x\| = |\alpha| \|x\|$
where x and y are vectors in vector space V and $\alpha \in \mathbb{R}$.

Basic concepts - norm

We will see the following norm in this course:

1) L_2 norm: for a vector x, the L_2 norm is defined as:

$$\|x\|_2 \equiv \sqrt{\sum_{i=1}^{n} x_i^2}$$

2) L_1norm: for a vector x,the L_2 norm is defined as:

$$\|x\|_1 \equiv \sum_{i=1}^{n} |x_i|$$

3) L_∞ norm: for a vector x,the L_∞ norm is defined as:

$$\|x\|_\infty \equiv \max_{i=1,\cdots,n} |x_i|$$

The norm for a matrix is induced as:

$$||A||_\alpha = \sup_{x \neq 0} \frac{||Ax||_\alpha}{||x||_\alpha}$$

Basic concepts - linear mapping, range and null space

We say a a map $x \to Ax$ is linear if for any $x, y \in \mathbb{R}^n$, and any $\alpha \in \mathbb{R}$,

$$A(x+y) = Ax + Ay$$
$$A(\alpha x) = \alpha Ax$$

It can be proved that each linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be expressed by the multiplication of a $m \times n$ matrix.
The range of linear operator $A \in \mathbb{R}^{m \times n}$, is the space spanned by the columns of A,

$$range(A) = \{y | \text{such that } y = Ax, x \in \mathbb{R}^n\}$$

The null space of linear operator $A \in \mathbb{R}^{m \times n}$ is the space,

$$null(A) = \{x | \text{such that } Ax = 0\}$$

It is "obvious" that $range(A)$ is perpendicular to $null(A^T)$. (exercise)

Basic concepts - four ways matrix multiplication

For the matrix-matrix product $B = AC$. If A is $l \times m$ and C is $m \times n$, then B is $l \times n$.
matrix multiplication method 1:

$$b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj}$$

Here b_{ij}, a_{ik}, and c_{kj} are entries of B, A, C.

Basic concepts - four ways matrix multiplication

For the matrix-matrix product $B = AC$. If A is $l \times m$ and C is $m \times n$, then B is $l \times n$.
matrix multiplication method 2:

$$B = [b_1 | b_2 | \cdots | b_n]$$

Here b_i is the $i - th$ column of matrix B.
Then,

$$B = [Ac_1 | Ac_2 | \cdots | Ac_n]$$
$$b_i = Ac_i$$

Each column of B is in the **range** (we will talk about it later) of A. Thus, the range of B is the subset of the range of A.

Basic concepts - four ways matrix multiplication

For the matrix-matrix product $B = AC$. If A is $l \times m$ and C is $m \times n$, then B is $l \times n$.

matrix multiplication method 3:

$$B = \begin{bmatrix} \tilde{b}_1^T \\ \tilde{b}_2^T \\ \cdots \\ \tilde{b}_l^T \end{bmatrix}$$

Here $\tilde{b}_i$ is the $i - th$ row of matrix B.

Then,

$$B = \begin{bmatrix} \tilde{a}_1^T C \\ \tilde{a}_2^T C \\ \cdots \\ \tilde{a}_l^T C \end{bmatrix}$$

$$\tilde{b}_i^T = \tilde{a}_i^T C$$

This form is not commenly used.

Basic concepts - four ways matrix multiplication

For the matrix-matrix product $B = AC$. If A is $l \times m$ and C is $m \times n$, then B is $l \times n$.
matrix multiplication method 4:

$$B = \sum_{i,j=1,\cdots,m} a_i \tilde{c}_j^T$$

Where, a_i is the $i-th$ column of matrix A, and $\tilde{c}_j^T$ is the $j-th$ row of matrix C.
Each term $a_i \tilde{c}_j^T$ is a rank-one matrix.

Outline

- Basic Concepts on Linear Algbra
  - vector space
  - norm
  - linear mapping, range, null space
  - matrix multiplication
- Iterative Methods for Linear Optimization
  - normal equation
  - steepest descent
  - conjugate gradient
- Unconstrainted Nonlinear Optimization
  - Optimality condition
  - Search direction
  - Line search

Linear Optimization- normal equation

We solve a linear system having n unknowns and with $m > n$ equations. We want to find a vector $\mathbf{m} \in \mathbb{R}^n$ that satisfies,

$$\mathbf{Fm} = \mathbf{d}$$

where $\mathbf{d} \in \mathbb{R}^m$ and $\mathbf{F} \in \mathbb{R}^{m \times n}$.
Reformulate the problem:

$$\text{define residual } \mathbf{r} = \mathbf{d} - \mathbf{Fm}$$

$$\text{find } \mathbf{m} \text{ that minimize} \|\mathbf{r}\|_2 = \|\mathbf{Fm} - \mathbf{d}\|_2$$

It can be proved that, we can minimize the residual norm when $\mathbf{F}^*\mathbf{r} = 0$. This is equivalent to a $n \times n$ system,

$$\mathbf{F}^*\mathbf{Fm} = \mathbf{F}^*\mathbf{d}$$

which is the normal equation. We can solve norm equation using direction methods such at LU, QR, SVD, Cholesky decomposition, etc.

Linear Optimization-steepest descent method

For the unconstraint linear optimization problem:

$$\min\ J(\mathbf{m}) = \|\mathbf{Fm} - \mathbf{d}\|_2^2$$

To find the minimum of objective function $J(\mathbf{m})$ iteratively using steepest descent method, at the current point $\mathbf{m_k}$, we update the model by moving along the nagative direction of gradient,

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \nabla J(\mathbf{m}_k)$$
$$\nabla J(\mathbf{m}_k) = \mathbf{F}^*(\mathbf{Fm}_k - \mathbf{d})$$

The gradient can be evaluated exactly, and we have analytical formula for the optimal α.

Linear Optimization-conjugate gradient method

For the unconstraint linear optimization problem:

$$\min\ J(\mathbf{m}) = \|\mathbf{Fm} - \mathbf{d}\|_2^2$$

Starting from $\mathbf{m}_0$, we have a series of search direction $\Delta\mathbf{m}_i, i = 1, 2, \cdots, k$, and updated model iteratively,$\mathbf{m}_i = \mathbf{m}_{i-1} - \alpha_{i-1}\Delta\mathbf{m}_{i-1}$, $i = 1, \cdots, k$.
For the next search direction $\Delta\mathbf{m}_k$ in the space span$\{\Delta\mathbf{m}_0, \cdots, \Delta\mathbf{m}_{k-1}, \nabla J(\mathbf{m}_k)\}$,

$$\Delta\mathbf{m}_k = \sum_{i=0}^{k-1} c_i \Delta\mathbf{m}_i + c_k \nabla J(\mathbf{m}_k)$$

The "magic" is that for linear problem $c_0 = c_1 = \cdots = c_{k-2} = 0$. We ended up with Conjugate gradient method,

$$\Delta\mathbf{m}_k = c_{k-1}\Delta\mathbf{m}_{k-1} + c_k \nabla J(\mathbf{m}_k)$$
$$\alpha_k = \min\ J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k)$$
$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k$$

We are searching within the space span$\{\Delta\mathbf{m}_0, \cdots, \Delta\mathbf{m}_{k-1}, \nabla J(\mathbf{m}_k)\}$ in CG method, though looks like we are doing a plane search.

Outline

- Basic Concepts on Linear Algbra
  - vector space
  - norm
  - linear mapping, range, null space
  - matrix multiplication
- Iterative Methods for Linear Optimization
  - normal equation
  - steepest descent
  - conjugate gradient
- Unconstrainted Nonlinear Optimization
  - Optimality condition
  - Search direction
  - Line search

Unconstrainted Nonlinear Optimization-Optimality condition

For the unconstraint nonlinear optimization problem:

$$\text{minimize}_{\mathbf{m}} J(\mathbf{m})$$

where $J(\mathbf{m})$ is a real-valued function.
How should we determine if $\mathbf{m}^*$ is a local minimizer?

Theorem

(First order necessary conditions for a local minimum)

$$\nabla J(\mathbf{m}^*) = 0$$

Theorem

(Second order necessary conditions for a local minimum)

$$\mathbf{s}^* \nabla^2 J(\mathbf{m}^*) \mathbf{s} \geq 0, \forall \mathbf{s} \in \mathbb{R}^n$$

Unconstrainted Nonlinear Optimization-Search direction

For the unconstraint nonlinear optimization problem:

$$\text{minimize}\ _{\mathbf{m}} J(\mathbf{m})$$

Given a model point $\mathbf{m}_k$, we want to find a search direction $\Delta\mathbf{m_k}$, and a real number, such that $J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) < J(\mathbf{m}_k)$.
How do we choose the search direction $\Delta\mathbf{m_k}$?
1) Gradient based method,

$$J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) - J(\mathbf{m}_k) \approx \alpha_k \nabla J(\mathbf{m}_k)^T \Delta\mathbf{m}_k + O(\|\Delta\mathbf{m}_k\|_2^2)$$

Thus,

$$\Delta\mathbf{m}_k = -\nabla J(\mathbf{m}_k)$$

is a search direction. We can also use similar technique in CG method,

$$\Delta\mathbf{m}_k = -c_1 \nabla J(\mathbf{m}_k) + c_2 \Delta\mathbf{m}_{k-1}$$

where $c_1, c_2 \in \mathbb{R}$.

Unconstrainted Nonlinear Optimization-Search direction

For the unconstraint nonlinear optimization problem:

$$\text{minimize}\ _{\mathbf{m}} J(\mathbf{m})$$

Given a model point $\mathbf{m}_k$, we want to find a search direction $\Delta\mathbf{m_k}$, and a real number, such that $J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) < J(\mathbf{m}_k)$.

How do we choose the search direction $\Delta\mathbf{m_k}$?

1) Methods based on a local quadratic model,

$$J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) - J(\mathbf{m}_k) \approx \alpha_k \nabla J(\mathbf{m}_k)^T \Delta\mathbf{m}_k + \alpha_k^2 \frac{1}{2} \Delta\mathbf{m}_k^T \nabla^2 J(\mathbf{m}_k) \Delta\mathbf{m}_k$$

We solve the approximated problem,

$$\text{minimize}\ \psi(\mathbf{p}_k) \equiv \nabla J(\mathbf{m}_k)^T \mathbf{p}_k + \frac{1}{2} \mathbf{p}_k \nabla^2 J(\mathbf{m}_k) \mathbf{p}_k$$

$$\mathbf{p}_k = \alpha_k \Delta\mathbf{m}_k$$

The approximated problem is a linear system and can be solved exactly. Then, update the model,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{p}_k$$

Unconstrainted Nonlinear Optimization-Line search

For the unconstraint nonlinear optimization problem:

$$\text{minimize }_{\mathbf{m}} J(\mathbf{m})$$

Given a model point $\mathbf{m}_k$, we want to find a search direction $\Delta\mathbf{m_k}$, and a real number, such that $J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) < J(\mathbf{m}_k)$.
How do we choose α_k for a given search direction $\Delta\mathbf{m_k}$? Can we choose arbitrary α_k such that $J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) < J(\mathbf{m}_k)$?
The answer is no. For example, $J(\mathbf{m}) = \mathbf{m}^2, \mathbf{m} \in \mathbb{R}^1$. We can find a sequence, such that

$$\mathbf{m}_0 = 2, \Delta\mathbf{m}_k = -\mathbf{m}_k$$

$$\alpha_k = \frac{2 + 3 \times 2^{-(k+1)}}{1 + 2^{-k}}$$

Then,

$$\mathbf{m}_k = (-1)^k (1 + 2^{-k})$$

$$J(\mathbf{m}_k) = \frac{1}{(1 + 2^{-k})^2} \to 1$$

Unconstrainted Nonlinear Optimization-Line search

For the unconstraint nonlinear optimization problem:

$$\text{minimize}_{\mathbf{m}} J(\mathbf{m})$$

Given a model point $\mathbf{m}_k$, we want to find a search direction $\Delta\mathbf{m_k}$, and a real number, such that $J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) < J(\mathbf{m}_k)$.
How do we choose α_k for a given search direction $\Delta\mathbf{m_k}$? A popular set of conditions that guarentee convergence named Wolfe condition:

$$J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k) \leq J(\mathbf{m}_k) + c_1 \alpha_k \nabla J(\mathbf{m}_k)^T \Delta\mathbf{m}_k$$

$$\nabla J(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k)^T \Delta\mathbf{m}_k \geq c_2 \alpha_k \nabla J(\mathbf{m}_k)^T \Delta\mathbf{m}_k$$

where $0 < c_1 < c_2 < 1$.

Reference

- Numerical Linear Algebra, by Lloyd N. Trefethen, David Bau III.
- Numerical Optimization, by Jorge Nocedal, Stephen Wright.
- Lecture notes from Prof. Walter Murray, http://web.stanford.edu/class/cme304/