TANA15/Lecture 7 - Contents

Non-Linear Least Squares

- Existance and Uniqueness results.
- The Newton and Gauss-Newton Methods.

Applications

- Application: Data Assimilation of Temperature measurements.

The Singular Value Decomposition

- Definition. Basic Properties.
- Linear Systems of Equations.
- The Condition Number.

Nonlinear Least Squares

Definition Let $r : \mathbb{R}^n \mapsto \mathbb{R}^m$, $m > n$, be a vector valued function. The *Least Squares Problem* is:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|r(x)\|_2, \qquad \frac{1}{2}\|r(x)\|_2^2 = \frac{1}{2}r(x)^T r(x) = \frac{1}{2}\sum_{i=1}^{m} r_i^2(x).$$

Questions Existance and Uniqueness of Solution? Algorithms?

Example With $r(x) = b - Ax$ we get the *linear least squares problem*.

Existance and Uniqueness

Consider the minimzation problem $\min f(x), f(x) : \mathbb{R}^n \mapsto \mathbb{R}$.

Definition A continuous function $f(x)$ is *coercive* if

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

Lemma A *coercive* function $f : \mathbb{R}^n \mapsto \mathbb{R}$ has a *global minimum*.

Example $f_1(x) = \mathrm{e}^{-x}$, $f_2(x) = x_1^2 - 2x_1x_2 + 3x_2^2$ and $f_3(x) = x_1^2 + x_1\mathrm{e}^{-x_2}$.

Definition A set S is *convex* if, $\alpha x + (1-\alpha)y \in S$, for any x, y in S, and $0 \leq \alpha \leq 1$.

Definition A function $f(x)$ is convex if, for any x, y,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \quad 0 \leq \alpha \leq 1.$$

If the inequality is *strict* then the $f(x)$ is *strictly convex*.

Lemma A strictly convex function $f(x)$ defined on a bounded convex set S has a unique global minimum.

Optimality conditions

Definition Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice differentiable The *Hessian matrix* of $f(x)$ is,

$$(H_f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad 1 \leq i,j \leq n.$$

Example The Taylor series is

$$f(x) = f(x^{(k)}) + (\nabla f(x^{(k)}))^T (x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T H_f(x^{(k)})(x - x^{(k)}) + \dots$$

Remark Locally $f(x)$ is approximated by a quadratic polynomial.

Definition A point x^* such that $\nabla f(x^*) = 0$ is called a *critical point*.

Lemma If x^* is a critical point and the Hessian $H_f(x^*)$ is positive definite then x^* is a *local minimum*.

Newtons Method

Find a local minimum x^* of $f(x)$ by solving the non-linear system of equations,

$$\nabla f(x) = 0,$$

using the Newton Method.

Algorithm Given a starting approximation $x^{(0)} \approx x^*$ do

$$H_f(x^{(k)}) s^{(k)} = -\nabla f(x^{(k)}), \qquad x^{(k+1)} = x^{(k)} + s^{(k)}.$$

until convergence. Verify that $x^{(k)}$ is a *local minimum* by checking if $H_f(x^{(k)})$ is positive definite.

Remark Very good method provided that the Hessian matrix H_f is easy to compute.

Recall that the *Least Squares Problem* is:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|r(x)\|_2, \qquad \frac{1}{2}\|r(x)\|_2^2 = \frac{1}{2} r(x)^T r(x) = \frac{1}{2} \sum_{i=1}^{m} r_i^2(x).$$

where $r : \mathbb{R}^n \mapsto \mathbb{R}^m$, $m > n$, is a vector valued function.

Questions What is the corresponding gradient vector and Hessian matrix?

Lemma Let $f(x) = r(x)^T r(x)$, $r(x) : \mathbb{R}^n \mapsto \mathbb{R}^m$, then

$$\nabla f(x) = J_r^T(x) r(x),$$

and

$$H_f(x) = J_r^T J_r + \sum_{i=1}^{m} r_i(x) H_{r_i}(x),$$

where J_r is the Jacobian of r and H_{r_i} is the Hessian of the component function r_i.

Remark The Hessian H_f is usually very computationally demanding to compute. In the *Gauss-Newton* method we approximate

$$H_f(x) \approx J_r(x)^T J_r(x).$$

The Gauss-Newton Method

Minimize $f(x) = \frac{1}{2} r(x)^T r(x)$ by the following steps.

Algorithm Let $x^{(0)}$ be a starting approximation and iterate

for $k = 1, 2, 3, \ldots$

Minimize $\|J_r(x^{(k)}) s^{(k)} + r(x^{(k)})\|_2$.

Update $x^{(k+1)} = x^{(k)} + s^{(k)}$.

if $\|s^{(k)}\| < tol$ **then** stop.

end

Remark The Jacobian can be computed using finite differences.

Example Fit a nonlinear model function,

$$f(t, x) = x_1 e^{x_2 t},$$

to the data

t	0.0	1.0	2.0	3.0
y	2.0	0.7	0.3	0.1

using the Gauss-Newton method and the starting approximation $x = (1, 0)^T$.

What is $r(x)$ and $J_r(x)$?

The Guass-Newton method applied to the problem gives

k	$x_1^{(k)}$	$x_2^{(k)}$	$\|r(x^{(k)})\|$	$\|s^{(k)})\|$
0	1.0000	0.0000	1.55e+00	9.21e-01
1	1.6900	-0.6100	4.61e-01	9.21e-01
2	1.9751	-0.9305	8.56e-02	4.29e-01
3	1.9941	-1.0036	4.50e-02	7.55e-02
4	1.9950	-1.0093	4.47e-02	5.81e-03
5	1.9950	-1.0095	4.47e-02	1.79e-04
6	1.9950	-1.0095	4.47e-02	4.76e-06
7	1.9950	-1.0095	4.47e-02	1.25e-07
8	1.9950	-1.0095	4.47e-02	3.26e-09
9	1.9950	-1.0095	4.47e-02	8.53e-11
10	1.9950	-1.0095	4.47e-02	2.23e-12

Remark Relatively quick convergence. Sometimes tricky to get the Gauss-Newton method to converge.

Data Assimilation and Temperature Measurements

Observation The steady state temperature in a wall satisfies

$$(k(x)T_x)_x = 0, \quad \text{for}, \quad a < x < b,$$

where $k(x)$ is the thermal conductivity.

If the wall consists of two different materials the function $k(x)$ may be given by,

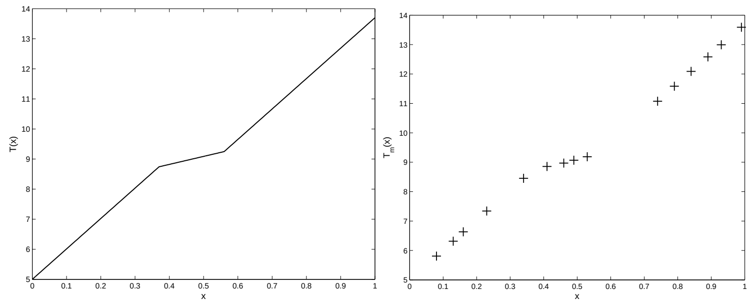
$$k(x) = \begin{cases} \kappa, & \alpha < x < \beta, \\ 1, & \text{otherwise}, \end{cases}$$

If, in additio, we have *boundary values* $T(a) = A$, $T(b) = B$ we can solve the problem and compute the steady state temperature

$$T(x) = f(u), \quad u = (A, B, \kappa, \alpha, \beta)^T.$$

Problem Suppose we measure the temperature at certain locations inside the wall can we still determine $T(x)$?

Example An exact temperature distribution $T_e(x)$ and $m = 15$ measurements in the interior of the wall.



How to formulate as a non-linear least squares problem?

Do the following:

1. Implement a numerical method that solves the boundary value problem given concrete values for the parameter vector u, i.e.

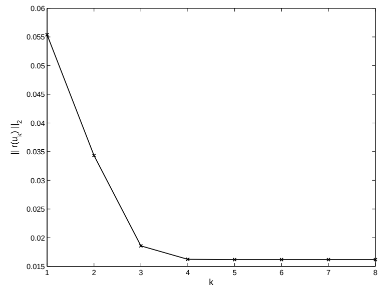
```
>> T = SteadyState( A,B,alpha,beta,kappa );
```

2. Let the measurement locations be given by $\{x_i\}_{i=1}^m$ and the measurement data be collected in a vector T_m.
3. Given a parameter vector $u^{(k)}$ we can compute the numerical solution $T(u^{(k)})$ and a residual vector,

$$r(u^{(k)}) = T_m - T(u^{(k)}),$$

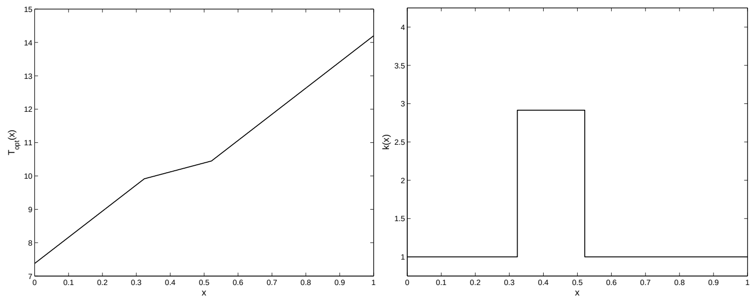
4. Compute the Jacobian J_r numerically using finite differences.

The convergence history $\|r(u^{(k)})\|_2$ during the Gauss Newton iterations.



Initial guess was $u^{(0)} = (7, 14, 0.28, 0.60, 3.02)^T$ and the final approximation $u^{(8)} = (7.38, 14.20, 0.32, 0.52, 2.9)^T$.

Remark A good initial guess is needed for convergence. The data points were constructed using $u_e = (7.4, 14.2, 0.32, 0.52, 3)^T$ and noise of size 10^{-2}.



The reconstructed temperature distribution $T(u^{(8)})$ and the corresponding thermal conductivity $k(x)$.

Remark This technique is called *Data assimilation*. All available equations are put into a non-linear least squares problem that is solved numerically.

The singular value decomposition

Proposition Every matrix $A \in \mathbb{R}^{m\times n}$ has a decomposition

$$A = U\Sigma V^T,$$

where U and V are orthogonal and $\Sigma \in \mathbb{R}^{m\times n}$ is *diagonal* with diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(n,m)} \geq 0$

Remark The diagonal elements $\{\sigma_i\}$ are called *singular values* and the columns $\{u_i\}$ of U and the columns $\{v_i\}$ of V are called right and left singular vectors.

Example Compute the SVD in Matlab by

```
>> A=[1 -2 3 ; -2 3 1 ; 2 -4 6 ; -1 2 -3];
>> [U,S,V]=svd(A); U , S
  U =
   -0.4025  0.0684  0.9129 -0.0000
    0.1675  0.9859 -0.0000  0.0000
   -0.8050  0.1368 -0.3651  0.4472
    0.4025 -0.0684  0.1826  0.8944

  S =
    9.2780       0        0
         0  3.4524        0
         0       0   0.0000
         0       0        0
```

Remark The matrix A has rank 2. V is 3×3 orthogonal.

Example Let $A \in \mathbb{R}^{4\times 3}$. Then

$$A = \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}^T.$$

Remark The vectors $\{u_i\}$ are a basis for $\mathbb{R}^4$ and the vectors $\{v_i\}$ are a basis for $\mathbb{R}^3$

A matrix $A \in \mathbb{R}^{m \times n}$ represents a *linear mapping*

$$A : \mathbb{R}^n \mapsto \mathbb{R}^m.$$

Remark If $U = (u_1, \ldots, u_m) \in \mathbb{R}^{m \times m}$ is *orthogonal* then the set of vectors $\{u_i\}$ form an orthogonal basis for $\mathbb{R}^m$.

Observation In the basis U, V the linear mapping is represented by the diagonal matrix D.

Lemma Let $A \in \mathbb{R}^{m \times n}$ and $A = U \Sigma V^T$. it holds that

$$A v_i = \sigma_i u_i \text{ and} A^T u_i = \sigma_i v_i, i = 1, 2, \ldots, \min(m, n).$$

Lemma Let $A \in \mathbb{R}^{m \times n}$ and $A = U \Sigma V^T$. We can write

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T.$$

Linear Systems of Equations

Lemma Let $A \in \mathbb{R}^{n \times n}$ be non-singular and $A = U \Sigma V^T$. Then the solution to the linear system $Ax = b$ is given by

$$x = V \Sigma^{-1} U^T b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.$$

Remarks The solution exists, i.e. A is non-singular, if $\sigma_n > 0$. If σ_n is very small the system is *Ill–conditioned*.

More expensive compared to using the LU factorization. Reveals linear dependencies among the columns of A.

Norms and the Condition Number

Recall If U is orthogonal and x is a vector then $\|Ux\|_2 = \|x\|_2$.

Lemma The norm is $\|A\|_2 = \sigma_1$.

Corollary The condition number is $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$.

Remark Previously we used $\|A\|_2 = (\lambda_{\max}(A^T A))^{1/2}$. Since $A^T A = U \Sigma^T \Sigma U^T$ we get $\lambda_i(A^T A) = \sigma_i^2$.