ECE240 Introduction to Linear Dynamical Systems
Lecture 9: Nonlinear Least-squares

Dr. Yu Zhang
ECE Department, UC Santa Cruz

Fall 2025

Outline

Nonlinear Least Squares: Problem Formulation

We seek a parameter vector $\mathbf{x} \in \mathbb{R}^n$ that best fits

$$\mathbf{r}(\mathbf{x}) := \begin{bmatrix} r_1(\mathbf{x}) \\ \vdots \\ r_m(\mathbf{x}) \end{bmatrix}.$$

Nonlinear Least Squares (NLLS) problem:

$$\min_{\mathbf{x}} \; f(\mathbf{x}) := \tfrac{1}{2}\|\mathbf{r}(\mathbf{x})\|_2^2 = \tfrac{1}{2}\sum_{i=1}^{m} r_i(\mathbf{x})^2.$$

Gradient: $\nabla f(\mathbf{x}) = \mathbf{J}(\mathbf{x})^\top \mathbf{r}(\mathbf{x})$, where $\mathbf{J}(\mathbf{x}) := \dfrac{\partial \mathbf{r}}{\partial \mathbf{x}}$.

Hessian:

$$\nabla^2 f(\mathbf{x}) = \mathbf{J}^\top \mathbf{J} + \sum_{i=1}^{m} r_i(\mathbf{x})\, \nabla^2 r_i(\mathbf{x}).$$

Examples of Nonlinear Least Squares

1. **Exponential model fitting:**

$$y_i \approx a e^{b t_i}, \quad r_i(\mathbf{x}) = a e^{b t_i} - y_i, \ \mathbf{x} = [a \ b]^\mathsf{T}.$$

2. **Logistic curve:**

$$y_i \approx \frac{L}{1 + e^{-k(t_i - t_0)}}, \qquad \mathbf{x} = [L \ k \ t_0]^\mathsf{T}.$$

3. **Robot calibration:**

$$r_i(\mathbf{x}) = f_{\mathrm{FK}}(\mathbf{x}; u_i) - y_i.$$

4. **Power system estimation:**

$$r_i(\mathbf{x}) = P_i^{\mathrm{model}}(\mathbf{x}) - P_i^{\mathrm{measured}}.$$

Many engineering estimation tasks reduce to NLLS.

Gauss-Newton Algorithm

Linearization at iteration k:

$$\mathbf{r}(\mathbf{x}_k + \Delta\mathbf{x}) \approx \mathbf{r}(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)\Delta\mathbf{x}.$$

Solve LS subproblem:

$$\min_{\Delta\mathbf{x}} \|\mathbf{J}_k\Delta\mathbf{x} + \mathbf{r}_k\|_2^2.$$

Normal equations:

$$\mathbf{J}_k^\top\mathbf{J}_k\Delta\mathbf{x} = -\mathbf{J}_k^\top\mathbf{r}_k.$$

Update:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}.$$

Convergence: If $\mathbf{J}(\mathbf{x}^*)$ has full column rank and residuals are small, Gauss-Newton converges quadratically.

Levenberg-Marquardt Algorithm (Damped Gauss-Newton)

Problem: Gauss-Newton may fail when $\mathbf{J}^\top \mathbf{J}$ is ill-conditioned or residuals are large.

LM step:

$$(\mathbf{J}_k^\top \mathbf{J}_k + \lambda_k \mathbf{I})\Delta \mathbf{x} = -\mathbf{J}_k^\top \mathbf{r}_k.$$

Update:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta \mathbf{x}.$$

Convergence:

- Global convergence with trust-region rule.
- Recovers GN's fast local convergence near solution.

Interpretation of LM Step

LM step:

$$(\mathbf{J}^\mathsf{T}\mathbf{J} + \lambda_k \mathbf{I})\Delta\mathbf{x} = -\mathbf{J}^\mathsf{T}\mathbf{r}.$$

When λ_k is very large:

$$\mathbf{J}^\mathsf{T}\mathbf{J} + \lambda_k \mathbf{I} \approx \lambda_k \mathbf{I}.$$

Thus, the step becomes

$$\Delta\mathbf{x} \approx -\frac{1}{\lambda_k}\mathbf{J}^\mathsf{T}\mathbf{r} = -\frac{1}{\lambda_k}\nabla f(\mathbf{x}_k),$$

which is exactly a gradient descent update with step size $\alpha = 1/\lambda_k$.

Summary:

- Small λ_k: behaves like Gauss-Newton (fast, curvature-aware).
- Large λ_k: behaves like gradient descent (safe, slower, robust).
- LM transitions smoothly between these regimes depending on λ_k.

Connection to Linear Least Squares

Linear LS is a special case of NLLS.

If residuals are linear: $\mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{y}$, then $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$.

Jacobian:

$$\mathbf{J}(\mathbf{x}) = \mathbf{A}.$$

Gauss-Newton step:

$$\mathbf{A}^\top \mathbf{A} \Delta\mathbf{x} = -\mathbf{A}^\top (\mathbf{A}\mathbf{x}_k - \mathbf{y}).$$

One-step convergence (assume $\mathbf{A}$ is full column rank):

$$\Delta\mathbf{x} = \mathbf{A}^\dagger \mathbf{y} - \mathbf{x}_k \quad \Rightarrow \quad \mathbf{x}_{k+1} = \mathbf{A}^\dagger \mathbf{y}.$$

Thus, linear LS is exactly Gauss-Newton in one iteration.

Convergence and Conditions

Local convergence (Gauss-Newton / LM):

- If $\mathbf{J}(\mathbf{x}^*)$ has full column rank and $\mathbf{r}(\mathbf{x}^*)$ is small, convergence is quadratic.
- With large residuals: linear rate.

Global convergence:

- LM (trust region) guarantees global convergence to stationarity.
- NLLS is generally nonconvex $\Rightarrow$ multiple local minima.

Ill-conditioning:

- If $\mathbf{J}^\top\mathbf{J}$ is ill-conditioned, LM/regularization helps.
- Scaling of parameters is important.

Example: Exponential Fitting with Gauss-Newton

Model:

$$y_i \approx a e^{b t_i}, \qquad \mathbf{x} = [a \ b]^{\mathsf{T}}.$$

Residuals:

$$r_i(\mathbf{x}) = a e^{b t_i} - y_i.$$

Jacobian:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} e^{b t_1} & a t_1 e^{b t_1} \\ e^{b t_2} & a t_2 e^{b t_2} \\ \vdots & \vdots \\ e^{b t_m} & a t_m e^{b t_m} \end{bmatrix}.$$

Gauss-Newton step:

$$\Delta \mathbf{x} = -(\mathbf{J}^{\mathsf{T}} \mathbf{J})^{-1} \mathbf{J}^{\mathsf{T}} \mathbf{r}.$$

Iterate until convergence.

Summary

- NLLS minimizes $\frac{1}{2}\|\mathbf{r}(\mathbf{x})\|_2^2$ with nonlinear residuals.
- Gauss-Newton uses linearization; LM adds damping for robustness.
- Linear LS is a special case with one-step convergence.
- Convergence: fast if residuals small and Jacobian well-conditioned.
- Widely used in data fitting, estimation, and inverse problems.