Least Squares Problem - 2

Review: Least square problems-1

- Linear least square
  - Cholesky decomposition
  - QR decomposition
  - SVD
- Non-linear least squares
  - Gauss Newton
  - Levenberg Marquardt(LM)
- Large Residue Problems

THE LEVENBERG–MARQUARDT METHOD

The Levenberg–Marquardt method is a Gauss-Newton method with a trust-region strategy. The use of a trust region avoids one of the weaknesses of Gauss– Newton, when the Jacobian J(x) is rank-deficient. The subproblem is

$$\min_p \frac{1}{2}||J_k p + r_k||_2^2$$
$$s.t.||p|| \leq \Delta_k$$

with model function

$$m_k(p) = \frac{1}{2}||r_k||^2 + p^T J_k^T r_k + \frac{1}{2} p^T J_k^T J_k p$$

When the solution p^{GN} of the Gauss-Newton of the above equation lies strictly inside the trust region (ie. $p^{GN} < \Delta_k$), then this step p^{GN} also solves the subproblem. Otherwise there is a $\lambda > 0$ such that the solution is p^{LM} that satisfies $||p|| = \Delta$ and

$$(J^T J + \lambda I)p = -J^T r$$

Disadvantages-

- Getting the best λ is a challenge
- May not work better for large residue problem (since the second factor $r^T\nabla_2 r$ is neglected in the optimization function)

LARGE RESIDUE PROBLEMS

Some times the residue of the model may be large due to the following reasons.

- Fault in the model (since we approximate a quadratic model or because we neglect the second factor $r^T\nabla_2 r$)
- Outliers in the data

If it is a large residue problem, Quasi-Newton converges much better. But in general we dont even know whether it is a large residue or small residue problem apriory. So we devise hybrid algorithms which behaves as GN or LM when the residue is small and switch to Newton or Quasi Newton when the residue is large.

Algorithm 1: First approach of a hybrid algorithm

1. Start as Quasi Newton: $\nabla^2 f = J^T J + r^T \nabla^2 r$
2. Find B_k for the current iterate (Use BFGS)
3. Try Gauss-Newton/Levenberg Marquardt direction $p_k^{GN/LM}$, and find x^{k+1}
4. If $\frac{f(x_{k+1})}{f(x_k)} < 5$, choose $p_k = p_k^{GN/LM}$
5. Overwrite B_k as $J_k^T J_k$
6. Else continue with Quasi-Newton, by computing $p_k = p_k^{QN}, x_{k+1}$
7. Repeat from step 2 till converges.

A second approach (Dennis, Gay, and Welsch)

By combining Gauss–Newton and quasi-Newton ideas, maintain approximations to just the second-order part of the Hessian and S_{k+1} using rank 2 update.

$$B_k = J^T J + S_k$$

This is probably the best-known algorithm in this class because of its implementation in the well-known NL2SOL package.

Constrained Optimization

General form:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subjected to} \quad \begin{cases} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases}$$

Where f is the objective function, while $c_i, i \in \mathcal{E}$ are the equality constraints and $c_i, i \in \mathcal{I}$are the inequality constraints. The feasible set,

$$\Omega = \{x | c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}$$

In short this can be written as $\min\limits_{x\in\Omega} f(x)$

A vector x^* is a local solution of the problem if $x^* \in \Omega$ and there is a neighbourhood N of x^* such that $f(x) \geq f(x^*)$ for $x \in N \cap \Omega$.

A vector x^* is a strict local solution(also called a strong local solution) if $x^* \in \Omega$ and there is a neighbourhood N of x^* such that $f(x) > f(x^*)$ for all $x \in N \cap \Omega$ with $x \neq x^*$.

A vector x^* is an isolated local solution if $x^* \in \Omega$ and there is a neighbourhood N of x^* such that x^* is the only local minimizer in $N \cap \Omega$.

An active set at any feasible point is the union of all equality constraints and the inequality constraints that are active at the point. Thus,
Active set $A(x) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(x) = 0\}$, where $x \in \Omega$

Example-1: Single equality constraint

$$\min_{x_1,x_2} x_1 + x_2 \quad \text{subjected to} \quad x_1^2 + x_2^2 - 2 = 0$$

Solution:

$$f(x) = x_1 + x_2 \qquad\qquad \mathcal{I} = \phi$$
$$c_1(x) = x_1^2 + x_2^2 - 2 \qquad\qquad \mathcal{E} = \{1\}$$

Figure 1: Example 1

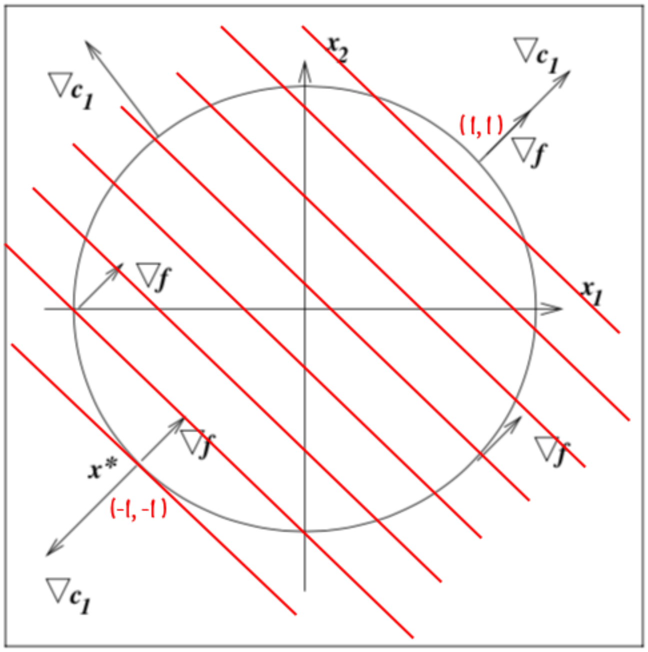


Figure 1 shows the optimal point, direction of the gradients of function and the constraints and also the contours of the constraints. At the solution point ∇f and ∇C are parallel. Hence we can write $\nabla f = \lambda \nabla C$.

<u>Proof for First Order Necessary Conditions.</u>

$c_1(x) = 0$ at any x satisfying the feasible set. Also $c_1(x + d) = 0$ to retain feasibility. Now using Taylor's expansion

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d$$

$$\Rightarrow \quad \nabla c_1(x)^T d = 0 \tag{1}$$

Similarly, a direction of improve must produce a decrese in f, so that

$$0 > f(x + d) - f(x) \approx \nabla f(x)^T d$$

$$\Rightarrow \quad \nabla f(x)^T d < 0 \tag{2}$$

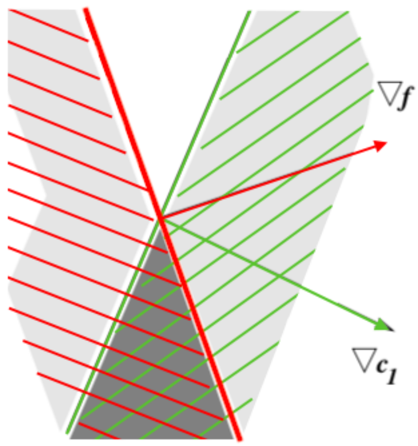
If no direction satisfy equations (1) and (2), the necessary condition for optimality is satisfied. As evident from the figure, this is possible iff ∇f and ∇C are parallel at the solution point. Hence we can write $\nabla f = \lambda \nabla C$.

By introducing the Lagrangian function

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x), \tag{3}$$
$$\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x) \tag{4}$$

At the solution x^* there is a scalar λ^* such that $\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0$.

Figure 2: Feasible region



This observation suggests that we can search for solutions of the equality-constrained problem by searching for stationary points of the Lagrangian function. The scalar quantity λ_1 is called a Lagrange multiplier for the constraint $c_1(x) = 0$. The condition $\nabla f = \lambda \nabla C$ is not a sufficient condition. For instance, in Example-1, equation (4) is satisfied at the point $x = (1, 1)$ with $\lambda = 0.5$, but this point is obviously not a solution—in fact, it maximizes the function $f(x)$ on the circle. Moreover, in the case of equality-constrained problems, we cannot turn the condition into a sufficient condition simply by placing some restriction on the sign of λ_1.

Example-2: Single inequality constraint

$$\min_{x_1, x_2} x_1 + x_2 \quad \text{subjected to} \quad 2 - x_1^2 - x_2^2 \geq 0$$

Solution:

$$f(x) = x_1 + x_2 \qquad\qquad \mathcal{E} = \phi$$
$$c_1(x) = 2 - x_1^2 - x_2^2 \geq 0 \qquad\qquad \mathcal{I} = \{1\}$$

Here the feasible region include the interior of the circle also. By inspection, we see that the solution is still $(-1, -1)$ and that the condition $\nabla f(x^*) = \lambda_1^* \nabla C(x^*)$ holds for the value

$\lambda_1^* = 0.5$. However, this inequality constrained problem differs from the previous equality constrained problem of Example-1 in that the sign of the Lagrangian multiplier plays a significant role.

The function improves if there is a direction d if $\nabla f(x)^T d < 0$

The feasibility is retained if $0 \leq c_1(x+d) \approxeq c_1(x) + \nabla c_1(x)^T d$

$$\Rightarrow \quad c_1(x) + \nabla c_1(x)^T d \geq 0 \tag{5}$$

Consider case-1, when x lies strictly inside the circle, so that $c_1(x) > 0$. Any vector d satisfies equation (5) provided its length is sufficiently small. This says that a direction d does not exist iff $\nabla f(x) = 0$.

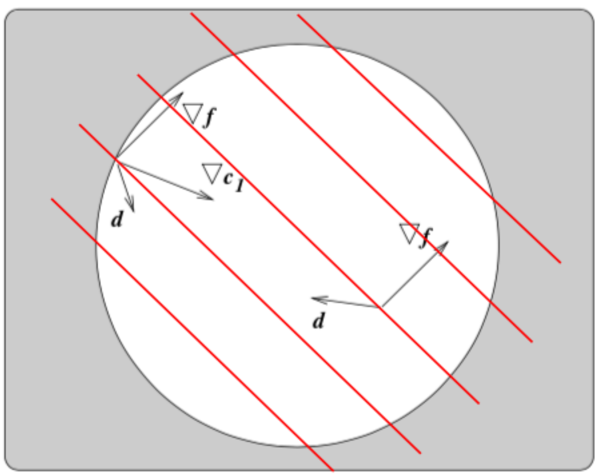
Case-2, when x lies in the boundary of the circle, so that $c_1(x) = 0$. Now the conditions are

$$\nabla f(x)^T d < 0 \tag{6}$$

$$\nabla c_1(x)^T d \geq 0 \tag{7}$$

Improvement directions from two different feasible points for the problem at which the constraint is active and inactive are shown in the Figure 3.

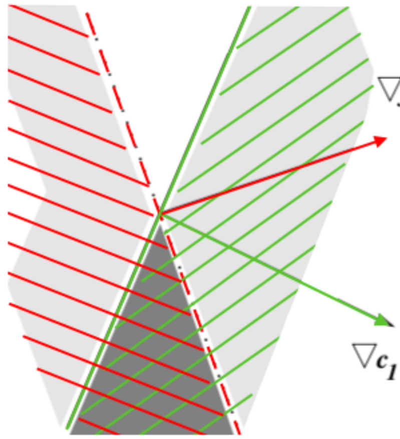
Figure 3: Improvement directions



The first of these conditions defines an open half-space, while the second defines a closed half-space, as illustrated in Figure 4. The two regions fail to intersect only when $\nabla f(x)$ and $\nabla c_1(x)$ are in the same direction.The sign of λ_1 is significant here.

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \quad \text{for some} \quad \lambda_1 \geq 0 \tag{8}$$

Figure 4: Example 2: Feasible Region



The optimality condition for both the cases 1 and 2 can be summarized through the Lagrangian function as,

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0 \quad \text{for some} \quad \lambda_1^* \geq 0 \tag{9}$$

$$\lambda_1^* c_1(x^*) = 0 \tag{10}$$

Equation (10) is called the complementary condition. It implies that the Lagrangian multiplier λ_1 can be positive only when the corresponding constraint c_1 is active. If $c_1(x*) \neq 0$ requires $\lambda_1^* = 0$ and thus the constraint is not active.