Methods for constrained problems

Niclas Börlin

5DA001 Non-linear Optimization

Methods for constrained non-linear problems

- Method for constrained non-linear problems may be categorized as follows:
  - Primal methods that work with the primal variables (x_i).
  - Dual methods that work with the dual variables (λ_i).
  - Combined methods that work on both.

  - Methods that use the original objective function.
  - Methods that modify the objective function.

Primal methods

- Primal methods that use the original objective function may be divided into two categories:
  - Feasible point methods, where every iterate x_i is a feasible point.
  - Penalty methods, that allow infeasible iterates, but the limit $\lim x_i = x^*$ is feasible.

Feasible point methods, linear equality constraints

- If our problem has linear equality constraints only

$$\begin{aligned} \min_{p \in \Re^n} \quad & f(\bar{x} + p) \\ \text{s.t.} \quad & Ap = 0 \end{aligned}$$

  we may solve

$$\min_{v \in \Re^{n-m}} \phi(v) = f(\bar{x} + Zv),$$

  instead, where Z is a null space matrix of A.
- The search direction p is found by first solving the reduced Newton equation (null-space equation)

$$Z^T \nabla^2 f(x) Z v = Z^T \nabla f(x)$$

  for v and then calculating $p = Zv$.

Feasible point methods, linear inequality constraints
Active set methods

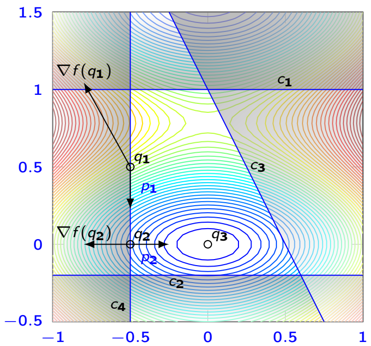
- For problems with linear inequality constraints there are e.g. active set methods.
- Active set methods solve for the minimum on a set $\mathcal{S}$ of active constraints and modifies the active set until the solution of the complete problem is found.

Feasible point methods, linear inequality constraints
Active set methods

- If x_k optimal on the current active set:
  - Calculate the Lagrange multipliers λ_i for all active constraints.
  - If the active set is empty or if all $\lambda_i \geq 0$,
    - Terminate. x_k is a local minimizer of the problem.
  - Otherwise, remove a constraint corresponding to a negative Lagrange multiplier λ_i from the active set.
- Determine a search direction p which is feasible with respect to the active constraints.
- Determine a step length α which satisfies $f(x_k + \alpha p) < f(x_k)$ and does not violate any inactive constraint.
- Update the point $x_{k+1} = x_k + \alpha p$.
- Modify the active set if any new constaints were activated.

Active set methods
Example

- At q_1,
  - $\mathcal{S} = \{c_4\}$, $\hat{A} = \begin{bmatrix} 2 & 0 \end{bmatrix}$, $Z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $p_1 = -ZZ^T\nabla f(q_1) \neq 0$.
  - q_1 is not minimizer on $\mathcal{S}$.
  - p_1 is descent direction on $\mathcal{S}$.
- At q_2,
  - $ZZ^T\nabla f(q_2) = 0$.
  - q_2 is minimizer on $\mathcal{S}$.
  - $\nabla f(q_2) = \hat{A}^T\lambda_4 \rightarrow \lambda_4 < 0$.
  - Remove c_4 from $\mathcal{S}$.
  - Compute p from new $\mathcal{S} = \emptyset$.
- At q_3,
  - $\nabla f(q_3) = 0$.
  - q_3 is minimizer on $\mathcal{S}$.



Feasible point methods, non-linear equality constraints

- Consider a non-linear problem with non-linear equality constraints:

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c(x) = 0. \end{aligned}$$

- If we construct the Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x),$$

we may apply Newton's method on the first order condition on $\mathcal{L}$, i.e.

$$\nabla\mathcal{L}(x, \lambda) = 0.$$

Feasible point methods, non-linear equality constraints

Cont'd

- The Newton formula for this problem is

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} p_k \\ \nu_k \end{bmatrix}.$$

- The vectors p_k and ν_k are found as the solution of the Newton equation for the Lagrangian function:

$$\nabla^2 \mathcal{L}(x_k, \lambda_k) \begin{bmatrix} p_k \\ \nu_k \end{bmatrix} = -\nabla \mathcal{L}(x_k, \lambda_k).$$

- Thus, for each iteration we calculate an update for both the primal parameters x and the dual parameters λ.

Sequential Quadratic Programming (SQP)

- The Newton equation for the Lagrangian function has the following structure:

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) & -\nabla c(x_k) \\ -\nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} p_k \\ \nu_k \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}(x_k, \lambda_k) \\ c(x_k) \end{bmatrix},$$

which corresponds to the first order condition for the problem

$$\begin{aligned} \min_p \quad & \tfrac{1}{2} p^T \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) p + p^T \nabla_x \mathcal{L}(x_k, \lambda_k), \\ \text{s.t.} \quad & \nabla c(x_k)^T p + c(x_k) = 0. \end{aligned}$$

- Thus, we use a second order Taylor approximation of the Lagrangian function and a first order Taylor approximation of the constaints.
- This technique is called Sequential Quadratic Programming (SQP) since we solve a sequence of quadratic problem.

Sequential Quadratic Programming (SQP)

Cont'd

- Derivation: The problem

$$\begin{aligned} \min_p \quad & \tfrac{1}{2} p^T \nabla^2_{xx}\mathcal{L} p + p^T \nabla_x \mathcal{L} \\ \text{s.t.} \quad & \nabla c^T p + c = 0 \end{aligned}$$

has the Lagrangian function

$$\mathcal{M}(p, \nu) = \frac{1}{2} p^T \nabla^2_{xx}\mathcal{L} p + p^T \nabla_x \mathcal{L} - (\nabla c^T p + c)^T \nu.$$

- The first order conditions are

$$\begin{aligned} \nabla_p \mathcal{M} &= \nabla^2_{xx}\mathcal{L} p + \nabla_x \mathcal{L} - \nabla c \nu = 0 \\ \nabla_\nu \mathcal{M} &= \nabla c^T p + c = 0 \end{aligned}$$

or

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L} & -\nabla c \\ -\nabla c^T & 0 \end{bmatrix} \begin{bmatrix} p \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L} \\ c \end{bmatrix}.$$

SQP for least squares problems

- For a least squares problem with non-linear equality constraints

$$\begin{aligned} \min_x \quad & f(x) = \frac{1}{2}\|r(x)\|^2 \\ \text{s.t.} \quad & c(x) = 0 \end{aligned}$$

we have

$$\nabla f = J^T r, \quad \nabla^2 f = J^T J + Q.$$

- The Lagrangian function is

$$\mathcal{L} = \frac{1}{2}\|r\|^2 - c^T \lambda$$

with partial derivatives

$$\nabla_x \mathcal{L} = J^T r - \nabla c \lambda, \quad \nabla_\lambda \mathcal{L} = -c,$$

$$\nabla^2_{xx}\mathcal{L} = J^T J + Q - \underbrace{\sum \lambda_i \nabla^2_{xx} c_i}_{Q_c}, \quad \nabla^2_{\lambda\lambda}\mathcal{L} = 0,$$

$$\nabla^2_{x\lambda}\mathcal{L} = -\nabla c, \quad \nabla^2_{\lambda x}\mathcal{L} = -\nabla c^T.$$

SQP for least squares problems
Cont'd

- Together

$$\nabla\mathcal{L} = \begin{bmatrix} J^T r - \nabla c\lambda \\ -c \end{bmatrix}, \quad \nabla^2\mathcal{L} = \begin{bmatrix} J^T J + Q - Q_c & -\nabla c \\ -\nabla c^T & 0 \end{bmatrix}.$$

- If we ignore the curvatures Q and Q_c we get the Gauss-Newton Equation for constrained problems

$$\begin{bmatrix} J^T J & -A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} p \\ \nu \end{bmatrix} = \begin{bmatrix} -J^T r + A^T\lambda \\ c \end{bmatrix},$$

where $A = \nabla c^T$.

SQP for least squares problems
Cont'd

- Geometrically, this corresponds to solving the problem

$$\begin{aligned} \min_p \quad & \frac{1}{2}\|Jp + r\|^2 \\ \text{s.t.} \quad & Ap + c = 0. \end{aligned}$$

- With substitution $w = -(Jp + r)$, the first row of the Gauss-Newton equation becomes

$$\begin{aligned} J^T Jp - A^T\nu &= -J^T r + A^T\lambda \\ &\Downarrow \\ -A^T\nu &= J^T w + A^T\lambda \\ &\text{or} \\ J^T w + A^T(\lambda + \nu) &= 0. \end{aligned}$$

- The vector $\lambda + \nu$ holds the updated Lagrange multiplier.

SQP for least squares problems
Cont'd

- If we denote the updated Lagrangian vector with $\lambda' = \lambda + \nu$, we may solve the following, equivalent, system equation:

$$\begin{bmatrix} 0 & 0 & A \\ 0 & I & J \\ A^T & J^T & 0 \end{bmatrix} \begin{bmatrix} \lambda' \\ w \\ p \end{bmatrix} = \begin{bmatrix} -c \\ -r \\ 0 \end{bmatrix}.$$

- Thus, with this formulation we do not need to estimate the Lagrangian multipliers. Instead we calculate an estimate for each iteration.

Obtaining global convergence
Merit functions

- In order to obtain a descent direction we approximate the (reduced) hessian with a positive definite matrix.
- To get an acceptable step length, we have to balance any reduction of the object function value with the violations of the constraints.
- One solution is to apply a line search on a merit function.
- An example of a quadratic merit function is

$$\mathcal{M}(x_k, \rho_k) = f(x_k) + \rho_k c(x_k)^T c(x_k) = f(x_k) + \rho_k \sum_{i=1}^{m} c_i(x_k)^2, \quad \rho_k > 0,$$

where the penalty parameter ρ controls the penalty term for violating the constraints.

Obtaining global convergence

Merit functions

- To obtain global convergence, the sequence $\{\rho_k\}$ must contain a non-decreasing sequence.
- Initially, the penalty value is low, the constraint is relaxed and the method is allowed to take short-cuts outside the feasible set.
- As the penalty parameter is increased, the iterate is forced to stay closer and closer to the constraint.
- The calculation of the penalty weights ρ_k is crucial to the efficiency of a merit-function based method.
- Increasing ρ too slowly will allow the iterates to stay unfeasible too long, increasing them too fast will force the iterates to follow the constraints unnecessarily close.

Penalty and barrier methods

- Consider the constrained minimization problem
$$\min f(x) \text{ s.t. } x \in S,$$
where S is the feasible set.
- Define
$$\sigma(x) = \begin{cases} 0 & x \in S; \\ \infty & \text{otherwise.} \end{cases}$$
- The constrained problem may thus be rewritten as an unconstrained problem
$$\min f(x) + \sigma(x).$$
- Since $\sigma(x) = \infty$ for all infeasible points, the minimum will be attained in a feasible point.
- Penalty and barrier methods formulate and solve a sequence of problems replacing $\sigma(x)$ with a continuous function that approaches $\sigma(x)$ as $k \to \infty$.
- Penalty methods impose a penalty for violating a constraint, barrier methods impose a penalty for getting too close to the constraint from the inside of the feasible set.

Barrier methods

- Barrier methods require the iterates to initially be feasible and approach the constraints from the inside.
- Barrier methods are suitable for inequality constrained problems.
- Consider the problem
$$\min f(x) \text{ s.t. } g_i(x) \geq 0, \; i = 1, \ldots, m.$$
- Define the function $\phi(x)$ such that $\phi(x) \to \infty$ as $g_i(x) \to 0$, e.g.
$$\phi(x) = -\sum_{i=1}^{m} \log(g_i(x))$$
or
$$\phi(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)}.$$
- This function will act as a barrier when x approaches the border of the feasible set from the inside.

Barrier methods

- The barrier function is defined as
$$\beta(x, \mu) = f(x) + \mu\phi(x),$$
where μ is called a barrier parameter.
- Barrier methods solve the sequence of problems
$$\min_x \beta(x, \mu_k)$$
for a sequence $\{\mu_k\}$ of positive barrier parameters that decrease monotonically to zero.
- As μ_k approaches zero, the penalty for being close to the constraint will decrease, and the point x will be allowed to come closer and closer to the constraints.
- With smaller μ_k, the barrier will become more and more vertical and the Hessian of the barrier function will become more and more ill conditioned.

Penalty methods

- Penalty methods allow infeasible iterates and are thus suitable also for equality constrained problems. Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) = 0, \quad i = 1, \ldots, m. \end{aligned}$$

- Define the function $\psi(x)$ such that $\psi(x) = 0$ as $x \in S$, otherwise $\psi(x) > 0$ and $\psi(x) \to 0$ as $x \to S$.
- The function $\psi(x)$ will impose a penalty that depends on how infeasible the point x is.
- An example of such a penalty is the quadratic-loss function

$$\psi(x) = \frac{1}{2} \sum_{i=1}^{m} g_i(x)^2 = \frac{1}{2} c(x)^T c(x).$$

Penalty methods

- The weight of the penalty is controlled by a positive penalty parameter ρ.
- As ρ increases, the function $\rho\psi$ approaches the ideal penalty σ.
- The penalty function is defined as

$$\pi(x, \rho) = f(x) + \rho\psi(x).$$

- Penalty methods solve a sequence of problems

$$\min_x \pi(x, \rho_k)$$

for an increasing sequence $\{\rho_k\}$.
- As $\rho_k \to \infty$, the iterates x_k will be forced closer and closer to the feasible set S.
- As with the barrier methods, for large ρ_k, the walls will be almost vertical and the Hessian of the penalty function will be ill conditioned.