“Fundamentals of Constraint Optimization"

Oscar S. Dalmau Cedeño
dalmau@cimat.mx

9 de octubre de 2018

Optimization Problem

***General Formulation*.**

$$\min\ f(x), \tag{1}$$
$$s.a:\ c_i(x) = 0 \qquad i \in \mathcal{E} \tag{2}$$
$$c_i(x) \geq 0 \qquad i \in \mathcal{I} \tag{3}$$

- We have studied the optimality conditions that characterize the solution of the previous problem.
- The next methods are iterative algorithms for finding local solutions.
- They generate a sequence of estimates of the solution x^*, that we expect to converge to an optimal solution.

Quadratic programming problems

$$\min\ q(x) = \frac{1}{2}x^T G x + x^T c, \tag{4}$$

$$s.a:\ a_i^T x = b_i \qquad i \in \mathcal{E} \tag{5}$$

$$a_i^T x \geq b_i \qquad i \in \mathcal{I} \tag{6}$$

- Solving quadratic programming problem is important for other algorithms: sequential quadratic programming methods and certain interior-point methods for nonlinear programming.

Penalty methods

Given the optimization problem with equality constraints

$$\min f(x), \tag{7}$$

$$s.a: \; c_i(x) = 0 \qquad i \in \mathcal{E} \tag{8}$$

Quadratic Penalty

$$x_k = \arg\min p(x, \mu_k) = f(x) + \mu_k \sum_{i \in \mathcal{E}} [c_i(x)]^2 \tag{9}$$

Penalty methods

- The penalty function is a combination of the objective function and the equality constraints.
- A solution of the original optimization problem can be obtained by solving a sequence of unconstrained problems $p(x, \mu_k)$.
- For example, for a problem with only equality constraints we can define the quadratic penalty function.
- $\mu_k > 0$ *the penalty parameter* is an increasing sequence, and we expect that the sequence $\{x_k\}$ converges to an optimal solution x^*.

Augmented Lagrangian methods

- The augmented Lagrangian function for equality-constrained is:

$$\mathcal{L}_{\mathcal{A}}(x,\lambda,\mu) \quad = \quad f(x) - \sum_{i\in\mathcal{E}} \lambda_i c_i(x) + \frac{\mu}{2}\sum_{i\in\mathcal{E}} [c_i(x)]^2$$

- The Augmented Lagrangian function combines the Lagrangian function and the quadratic penalty function.

Augmented Lagrangian methods

- The optimization algorithm is an iterative procedure
- First, these methods fix the parameter $\mu > 0$ and an estimate of λ.
- Then, find a value of x that approximately minimizes $\mathcal{L}_{\mathcal{A}}(\cdot, \lambda, \mu)$.
- After the previous step, the parameters $\mu > 0$ and λ are updated and the process is repeated.
- These methods avoids some limitations of the quadratic penalty function.

Sequential quadratic programming (SQP)

- In these methods a quadratic programming subproblem is solved at each iterate.
- This subproblem is formulated in terms of the search direction.
- The search direction p_k at the iterate (x_k, λ_k) is the solution of:

$$\text{min } \frac{1}{2} p^T \nabla^2_{xx} \mathcal{L}(x_k, \lambda_k) p + \nabla f(x_k)^T p$$

$$\begin{aligned} \text{s.t: } \nabla c_i(x_k)^T p + c_i(x_k) &= 0, i \in \mathcal{E}, \\ \nabla c_i(x_k)^T p + c_i(x_k) &\geq 0, i \in \mathcal{I} \end{aligned}$$

Sequential quadratic programming (SQP)

- The objective function in this subproblem is a quadratic approximation of the Lagrangian at $x_k + p$, ie, it can be proved that $\nabla_x \mathcal{L}(x_k, \lambda_k)^T p = \nabla f(x_k)^T p$
- The constraints are the linear approximations at $x_k + p$ of the original constraints

Elimination using linear constraints

Nonlinear function subject to a set of linear equality constraints,

$$\begin{aligned} \text{min } & f(x) \\ \text{s.t: } Ax &= b \end{aligned}$$

where A is a full row rank of size $m \times n$ matrix with $m \leq n$.

Elimination using linear constraints

- We can find a subset of m columns of A that is linearly independent.
- Let B be a matrix with these columns, ie, B is an invertible matrix of size $m \times m$.
- Let P be a permutation matrix of size $n \times n$, such that

$$AP = [B, N]$$

where N denotes the $n - m$ remaining columns of A.

Elimination using linear constraints

- Let us define the subvectors $x_B \in \mathbb{R}^m$ and $x_N \in \mathbb{R}^{n-m}$

$$\begin{bmatrix} x_B \\ x_N \end{bmatrix} = P^T x$$

where x_B are the *basic variables* and B the *basis matrix*.

- As $PP^T = I$

$$\begin{aligned} b &= Ax = (AP)(P^T x) = Bx_B + Nx_N \\ x_B &= B^{-1}(b - Nx_N) \end{aligned}$$

Elimination using linear constraints

Then

$$\begin{aligned} \text{min } & f(x) \\ \text{s.t: } Ax & = b \end{aligned}$$

can be written by substitution (elimination of variables) as follows

$$f(x) \; = \; f(PP^T x) = f\left(P\begin{bmatrix} x_B \\ x_N \end{bmatrix}\right) = f\left(P\begin{bmatrix} B^{-1}(b - Nx_N) \\ x_N \end{bmatrix}\right)$$

$$\text{min}_{x_N} h(x_N) \; := \; f\left(P\begin{bmatrix} B^{-1}(b - Nx_N) \\ x_N \end{bmatrix}\right)$$

Elimination using linear constraints

- $x_B \in \mathbb{R}^m$ and $x_N \in \mathbb{R}^{n-m}$

$$
\begin{aligned}
P^T x &= \begin{bmatrix} x_B \\ x_N \end{bmatrix} \\
&= \begin{bmatrix} B^{-1}b - B^{-1}Nx_N \\ 0 + Ix_N \end{bmatrix} \\
&= \begin{bmatrix} B^{-1} \\ 0 \end{bmatrix} b + \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix} x_N \\
&= Yb + Zx_N
\end{aligned}
$$

with $Y = \begin{bmatrix} B^{-1} \\ 0 \end{bmatrix}$ and $Z = \begin{bmatrix} -B^{-1}N \\ I \end{bmatrix}$

Elimination using linear constraints

- Z has $n - m$ linearly independent columns (because of the presence of the identity matrix in the lower block) and that it satisfies $(AP)Z = [B, N]Z = 0$.
- Z is a basis for the null space of A.
- The columns of Y and the columns of Z form a linearly independent set.

Elimination using linear constraints

Note from

$$[x_B; x_N] \quad = \quad Yb + Zx_N$$

- That Yb is a particular solution of the linear constraints $Ax = b$, ie, $x = [x_B; x_N]$ with $x_B = B^{-1}b$ and $x_N = 0$.
- The term Zx_N represents the displacement along the null space of the constraints (the null solution) , ie,

$$[B, N]Zx_N \quad = \quad 0x_N = 0$$

General elimination using orthogonal basis

- Let us choose matrices $Y \in \mathbb{R}^{n\times m}$ and $Z \in \mathbb{R}^{n\times(n-m)}$ such that

$$D = [Y,\ Z] \in \mathbb{R}^{n\times n}$$

  is nonsingular, ie, $rk(D) = n$, and $AZ = 0$.
- Then $rk(AD) = rk(A)$, see corollary 1 in the notes at the end, ie, $rk(AD) = m$.
- On the other hand, $AD = A[Y,\ Z] = [AY,\ AZ] = [AY,\ 0]$, then $rk(AY) = rk(AD) = m$ due to $rk(AD) = m$.
- As $AY \in \mathbb{R}^{m\times m}$ and $rk(AY) = m$ then AY es invertible!

General elimination using orthogonal basis

- The solution of equation $Ax = b$ can be written as

$$x = Yx_Y + Zx_Z$$

- The solution of equation $Ax = b$ can be written as

$$b = Ax = AYx_Y + AZx_Z = AYx_Y$$

then $x_Y = (AY)^{-1}b$

- Then

$$x = Y(AY)^{-1}b + Zx_Z$$

satisfies the constraints $Ax = b$ for any $x_Z \in \mathbb{R}^{n-m}$

General elimination using orthogonal basis

- Finally, the original optimization problem with equality constraints is equivalent to the following unconstraint optimization problem

$$\min_{x_Z} h(x_Z) \quad := \quad f(Y(AY)^{-1}b + Zx_Z)$$

General elimination using orthogonal basis

- $Y(AY)^{-1}b$ is a particular solution of the linear constraints $Ax = b$, ie, $Y(AY)^{-1}b$ is a solution for $x_Z = 0$.
- The term Zx_Z represents the displacement along the null space of the constraints (the null solution) , ie,

$$AZx_Z \;=\; 0x_Z = 0$$

General elimination using orthogonal basis

- How to select Y? The idea is to select Y such that AY is well-conditioned
- For that purpose, we can use the QR decomposition with column pivoting (see notes at the end), ie

$$A^T\Pi = QR = [Q_1, Q_2][R; 0]$$

  where Π is a permutation matrix, with $\Pi\Pi^T = I$.
- QR column pivoting is useful when A is (nearly) rank deficient.

General elimination using orthogonal basis

- Now, we define $Y = Q_1$ and $Z = Q_2$ that form an orthonormal basis in $\mathbb{R}^n$, and satisfy $Y^T Y = I$, $Z^T Z = I$ and $Y^T Z = 0$
- Therefore

$$
\begin{aligned}
A^T \Pi &= [Q_1, Q_2][R; 0] = YR \\
A^T &= YR\Pi^T \\
A &= \Pi R^T Y^T
\end{aligned}
$$

then

$$
\begin{aligned}
AY &= \Pi R^T \\
AZ &= 0
\end{aligned}
$$

General elimination using orthogonal basis

- Y and Z have the desired properties
- The condition number of AY is the same as that of R
- Therefore

$$
\begin{aligned}
x &= Y(AY)^{-1}b + Zx_Z \\
&= Q_1(\Pi R^T)^{-1}b + Q_2 x_Z \\
&= Q_1 R^{-T}\Pi^T b + Q_2 x_Z
\end{aligned}
$$

General elimination using orthogonal basis

- Note that

$$
\begin{aligned}
A^T(AA^T)^{-1} &= YR\Pi^T(\Pi R^T Y^T Y R \Pi^T)^{-1} \\
&= Q_1 R \Pi^T \Pi R^{-1} R^{-T} \Pi^T \\
&= Q_1 R^{-T} \Pi^T
\end{aligned}
$$

- Then

$$
A^T(AA^T)^{-1}b = Q_1 R^{-T} \Pi^T b
$$

is the particular solution

General elimination using orthogonal basis

- On the other hand, $A^T(AA^T)^{-1}$ is the solution of the minimum norm problem

$$\begin{aligned} \text{min } & \|x\| \\ \text{s.t: } Ax &= b \end{aligned}$$

- The elimination strategy using the orthogonal basis is a good method from numerical stability point of view.
- The QR factorization is the main computational cost of this reduction strategy.

Example 1

$$\begin{aligned} \text{min } f(x) &= f(x_1, x_2, x_3, x_4) \\ \text{s.t: } x_1 + x_3^2 - x_3 x_4 &= 0 \\ -x_2 + x_3^2 + x_4 &= 0 \end{aligned}$$

from the constraint we can compute x_1, x_2 and substitute in $f()$, ie

$$\begin{aligned} x_1 &= -x_3^2 + x_3 x_4 \\ x_2 &= x_3^2 + x_4 \end{aligned}$$

and obtain the equivalent problem

$$\text{min}_{x_3, x_4} h(x_3, x_4) \quad := \quad f(-x_3^2 + x_3 x_4, x_3^2 + x_4, x_3, x_4)$$

Example 2

$$\begin{aligned} \text{min } f(x) &= x_1^2 + x_2^2 \\ \text{s.t: } (x_1 - 1)^3 &= x_2^2 \end{aligned}$$

using the elimination strategy we obtain

$$\text{min}_{x_1} h(x_1) := x_1^2 + (x_1 - 1)^3$$

- However, this problem is not equivalent to the original, because the nonlinear constraint in the original problem has an implicit constraint, ie, $x_1 \geq 1$ which is not considered in the elimination process!!.
- If we wish to eliminate x_2, we should explicitly introduce the constraint $x_1 \geq 1$ in the new problem.

Comments

- The use of nonlinear equations to eliminate variables may result in errors that can be difficult to figure out.
- For this reason, nonlinear elimination is not used by most optimization algorithms.
- Many algorithms first linearize the constraints and apply elimination techniques to the simplified problem

Note: QR decomposition with column pivoting

QR decomposition with column pivoting introduces a permutation matrix P:

$$AP = QR \quad \iff A = QRP^T$$

Column pivoting is useful when A is (nearly) rank deficient, or is suspected of being so. P is usually chosen so that the diagonal elements of R are non-increasing:

$$|r_{11}| \geq |r_{22}| \geq \ldots \geq |r_{nn}|$$

Note: Wikipedia

Note on matrix rank

Theorem

If A and B are two matrices which can be multiplied, then rank(AB) <= min(rank(A), rank(B)).

Corollary 1

If A is an m by n matrix and B is a square matrix of rank n, then rank(AB) = rank(A).

Corollary 2

If A is an m by n matrix and B is an n by m matrix, and both are of rank m, then rank(AB) = m.

Note on matrix rank

Corollary 1

If A is an m by n matrix and B is a square matrix of rank n, then $rk(AB) = rk(A)$.

B is an n by n matrix of rank n, ie, $rk(B) = n$, then B has inverse and $rk(B^{-1}) = n$. Therefore

$$rk(A) = rk(ABB^{-1}) \leq \min\,(rk(AB), rk(B^{-1})) \leq rk(AB)$$

due to $\min\,(x, y) \leq x$.
On the other hand,

$$rk(AB) \leq \min\,(rk(A), rk(B)) \leq rk(A)$$

Then

$$rk(AB) = rk(A)$$

Note on matrix rank

Corollary 2

If A is an m by n matrix and B is an n by m matrix, and both are of rank m, then $rk(BA) = m$.

$rk(A) = rk(B) = rk(B^T) = rk((B^T B)^{-1}) = m$, on the other hand

$$rk(BA) \leq \min\,(rk(B), rk(A)) = m$$

and

$$\begin{aligned} m = rk(A) &= rk((B^T B)^{-1} B^T BA) \\ &\leq \min\,(rk((B^T B)^{-1}), rk(B^T), rk(BA)) \\ &\leq rk(BA) \end{aligned}$$

then $rk(BA) = m$