Principles of constrained optimization: equality constraints

C6-7 of " Calculus: Real analysis and optimization"

Professor Manuel Samuelides

ISAE, University of Toulouse
Toulouse, France

October 2011

Outline

Outline

Presentation of the QP problem

Definition

A QP problem ("quadratic programming") problem is the convex optimization problem

$$\begin{cases} \min_x J(x) \\ Cx = d \end{cases}$$

where $J(x) = \frac{1}{2}(x \mid Ax) + (b \mid x)$ *is a quadratic objective function defined on the configuration space* $\mathcal{E} = \mathbb{R}^n$ *and where* $\Gamma = \{x \in \mathcal{E} \textit{ such that } Cx = d\}$ *is the affine space of* ***feasible configurations****.*

Solution of the QP problem

- If A is definite positive, objective function J is strictly convex with compact level sets and Γ is a convex set. so we saw that the QP problem admits one and only one solution.
- C is a (p, n) matrix with $p \leq n$. If C is of rank p, the p constraints are independent. The admissible configuration space is of dimension $n - p$ and the constraints are said **qualified**
- If the constraints are qualified, it is possible to solve the QP problem by linearly parametrizing the feasible configuration subspace Γ. The QP problem amounts to the unconstrained minimization of a quadratic objective function on $\mathbb{R}^{n-p}$ and is solved by a linear system. We shall introduce more efficient methods

Outline

1. Quadratic programming
   - Notations and definitions
   - Lagrangian function
2. Applications of quadratic programming
   - Pseudo-inverse
   - Optimal control with quadratic cost in a linear model
3. Equality constraints
   - Lagrange solutions
4. Algorithmics
   - Newton Lagrange method (SQP)
   - Constraint penalization algorithm

Lagrange multipliers

Theorem

There exists $\lambda^* \in \mathbb{R}^p$ such that the solution x^* of QP checks
$Ax^* + b + C^t\lambda = 0 \Leftrightarrow Ax^* + b + \sum_{i=1}^p \lambda_i^* c_i^t = 0$

Proof

Let us choose any admissible line of parametric equation $\gamma(t) = x^* + tu \in \Gamma$. Actually we have $Cu = 0$ and u is an admissible direction. We have

$$J(\gamma(t)) = J(x^*) + t[(x^* \mid Au) + (b \mid u) + \frac{t^2}{2}(u \mid Au)$$

So $$Cu = 0 \Rightarrow (Ax^* + b \mid u) = 0$$

End of the proof of Lagrange multipliers theorem

We shall have

$$\forall i, (c_i^t \mid u) = 0 \Rightarrow (Ax^* + b \mid u) = 0$$

The proof relies on the bi-orthogonality lemma

Lemma

The biorthogonal of a set of vectors is the vector space generated by this set.

Then there exists λ_i^* such that $Ax + b + \sum_i \lambda_i^* c_i^t = 0$.
Moreover, if the constraints are qualified this solution is unique.

Definition

*The λ_i^* are called the **Lagrange multipliers** of the QP problem.*

Lagrangian optimality system

Theorem

Let (A, b, C, d) define a QP problem with strictly convex objective and qualified constraints. Then the following linear system

$$\begin{cases} Ax + b + C^t\lambda &= 0 \\ Cx &= d \end{cases}$$

admits one unique solution (x^*, λ^*) defined by

$$\begin{cases} x^* &= -A^{-1}(C^t\lambda + b) \\ \lambda^* &= -(CA^{-1}C^t)^{-1}(CA^{-1}b + d) \end{cases}$$

x^* is the unique solution of the QP problem and λ^* is the vector of associated Lagrange multipliers.

Proof of Lagrange optimality system solution

- The solution is clear since the invertibility of $CA^{-1}C^t$ is ensured by the hypothesis on the QP system.
- x^* and λ^* has to check the optimality system as it was seen previously. So they are equal to its unique solution.
- They were already found unique. Note that the strict convexity of the objective and the qualification of the constraints are crucial to ensure the characterization by the optimality system.

Introduction of Lagrangian function

Definition

The Lagrangian function of the QP problem (A, b, C, d) is the function L defined on $\mathcal{E} \times \mathbb{R}^p$ by

$$L(x, \lambda) = J(x) + (\lambda \mid Cx - d)$$

Proposition

The unique solution of the QP problem is the unique stationary point of its Lagrangian.

Proof

It's just another formulation of the Lagrangian optimality system characterization. This system is the stationarity system of the Lagrangian where its gradient is equal to 0.

Dual problem

Theorem

*Let the **dual function** $\mathcal{J}$ of the QP problem be defined on $\mathbb{R}^p$ by $\mathcal{J}(\lambda) = \min_{x\in\mathcal{E}} L(x, \lambda)$.*
$\mathcal{J}$ is a strictly concave quadratic function. It is maximum in λ^*

Proof We already compute the solution of the Lagrangian minimization in x: $x^* - A^{-1}(C^t\lambda + b)$
When we report it in the Lagrangian, we find

$$\mathcal{J}(\lambda) = \frac{-1}{2}(b + C^t\lambda \mid A^{-1}(b + C^t\lambda)) - (d \mid \lambda)$$

which is strictly concave in λ since we saw that $CA^{-1}C^t$ is definite positive. Checking the maximum of $\mathcal{J}$ is λ^* is easy.

Outline

Pseudo-inverse

Theorem

If C is a (n,p) matrix of full rank p (with $p \leq n$) the unique solution of the QP problem

$$\begin{cases} \min_{x\in\mathbb{R}^n} \| x \|^2 \\ Cx = d \end{cases}$$

is $x^* = C^t(CC^t)^{-1}d$

Proof

We just apply the previous theorems with $A = I$ and $b = 0$

Remark

The regularization of an ***ill-posed*** *problem by transforming it into an optimization problem is currently used.*

Penalization of the constraint

An alternative way of solving an ill-posed problem consists in minimizing an objective which is a compromise between the problem and the regularity of the solution.

$$\min_{x \in \mathbb{R}^n} \| x \|^2 + r \| Cx - d \|^2$$

The solution is $\hat{x}(r) = (\frac{1}{r} + C^t C)^{-1} C^t d$

- Notice that it is defined even when C is not full rank.
- This method has a bayesian interpetation as a MAP estimation of the state x with a prior Gaussian probability and a Gauss measurement noise $\epsilon \sim \mathcal{N}(0, \frac{1}{r})$ with $d = Cx + \epsilon$

Outline

Optimal control problem for a linear system

- Let us consider the control problem for the discrete-time controlled linear dynamic system

$$\forall t \in \{0, .., T-1\},\ x_{t+1} = A_t x_t + B_t u_t$$

- The open-loop control is the sequence $(u_0, ..., u_{T-1})$
- The dynamic equation will be the constraints;
- The objective function incorporates the true objective (final state, reference trajectory...) and the control cost.

Optimal control as a QP problem

We put here a final state problem on the form of a QP problem

$$\begin{cases} \min J(x,u) = \frac{1}{2}(x_T \mid Qx_T) + \frac{1}{2}\sum_{t=0}^{T-1}(u \mid Ru) \\ \forall t \in \{0,..,T-1\}\ x_{t+1} - A_t x_t - B_t u_t = 0 \end{cases}$$

The Lagrange optimality linear system is

$$\begin{cases} \forall t \in \{0,..,T-1\}, & x_{t+1} - A_t x_t - B_t u_t = 0 \\ & Qx_T + \lambda_T = 0 \\ \forall t \in \{1, T-1\}, & \lambda_t - A_t^t \lambda_{t+1} = 0 \\ \forall t \in \{0, T-1\}, & Ru_t - B_t^t \lambda_{t+1} = 0 \end{cases}$$

Solution of the optimal control problem

- Note that all the solution is determined by the knowledge of the final state x_T.
- From x_T,we determine the final **adjoint state** λ_T.
- We then determine all the adjoint states λ_t by back-propagation.
- We are then able to determine all the control sequence (u_t).
- Then we have to adjust the initial condition with the final state.

Outline

ENLP Problem

We call ENLP problem the following optimization problem with equality constraints

$$\begin{cases} \min_x J(x) \\ C(x) = 0 \end{cases}$$

- It is not useful to make convexity assumptions when C is non linear.
- We suppose that J is a C^2 function on $\mathcal{E} = \mathbb{R}^n$ with compact level sets,
- we suppose that C is a C^1 application of $\mathcal{E}$ in $\mathbb{R}^p$ which is of full rank in the sense of the rank of its gradient ∇C.

Satisfying the constraints

- We define the set $\Gamma = \{x \in \mathcal{E} \text{ such that } C(x) = 0$ of the **feasible configurations**
- If $x \in \Gamma$ we define from differentiable path γ such that $\gamma(0) = x$ and $\gamma(t) \in \mathcal{E}$ the space of **admissible directions** u which check $\nabla C_x u = 0$.
- It is possible using the compact level set property to show the existence of a solution for ENLP.

Lagrange optimality system: first order conditions

Theorem

Let x^* be a solution of ENLP. If the constraints are qualified in x^* then there exists a Lagrange multiplier vector $\lambda^* \in \mathbb{R}^p$ such that (x^*, λ^*) is a stationary point of the lagrangian $L(x, \lambda) = J(x) + (\lambda \mid C(x))$ which is equivalent to the following first order Lagrange conditions

$$\begin{cases} \nabla J(x^*) + \nabla C(x^*)^t \lambda = 0 \\ C(x^*) = 0 \end{cases}$$

Proof The proof is the same than for QP using the admissible directions and the biorthogonal lemma.

Lagrange optimality system: second order conditions

The first order conditions are necessary.
Since the convexity assumptions are missing, the first order conditions are no more sufficient even to ensure that the solution is a local minima.
One has to consider necessary second order conditions:

Proposition

Let $x^* \in \Gamma$ be a feasible configuration satisfying the first order conditions for a Lagrange multiplier vector. If any admissible direction u in x^* checks the following second-order condition

$$(u \mid \nabla^2 J(x^*)u) > 0$$

then x^* is a local minima of J in Γ.

Outline

1. Quadratic programming
   - Notations and definitions
   - Lagrangian function
2. Applications of quadratic programming
   - Pseudo-inverse
   - Optimal control with quadratic cost in a linear model
3. Equality constraints
   - Lagrange solutions
4. Algorithmics
   - Newton Lagrange method (SQP)
   - Constraint penalization algorithm

Newton method (SQP)

Formally the Newton Lagrange method is an algorithm which applies at each step a second-order approximation of the Lagrange optimality system.

Algorithm

At each step k with (x_k, λ_k) solve:

$$\begin{pmatrix} \nabla^2_{xx} L(x_k, \lambda_k) & \nabla C(x_k) \\ \nabla C(x_k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = - \begin{pmatrix} \nabla L(x_k, \lambda_k)^t \\ C(x_k) \end{pmatrix}$$

A local convergence theorem of this algorithm is proven in [1] in the neighbourhood of the stationary feasible points of the lagrangian under regularity assumptions. So it does not ensure a secure solution.

Outline

1. Quadratic programming
   - Notations and definitions
   - Lagrangian function
2. Applications of quadratic programming
   - Pseudo-inverse
   - Optimal control with quadratic cost in a linear model
3. Equality constraints
   - Lagrange solutions
4. Algorithmics
   - Newton Lagrange method (SQP)
   - Constraint penalization algorithm

Courant algorithm

We shall consider the method of increasing penalization of constraints to achieve unconstrained optimization problems

Theorem

For an ENLP problem set $\mathcal{V}(x,\sigma) = J(x) + \frac{\sigma}{2} \parallel C(x) \parallel^2$ *Let* $\sigma_n \to \infty, x_n = \arg\min_{x\in\mathcal{E}} \mathcal{V}(x,\sigma_n)$. *Then*

a) $\mathcal{V}(x_n,\sigma_n) \leq \mathcal{V}(x_{n+1},\sigma_{n+1})$

b) $\parallel C(x_n) \parallel \geq \parallel C(x_{n+1}) \parallel$ *croissante.*

c) $J(x_n) \leq J(x_{n+1})$

d) $\parallel C(x_n) \parallel \to 0$

e) *The limit points of the sequence* (x_n) *are solutions of ENPL. If ENPL solution is unique, then the algorithm converges towards this solution.*

Proof of Courant algorithm

(a) We have
$\mathcal{V}(x_n, \sigma_n) \leq \mathcal{V}(x_{n+1}, \sigma_n) \leq \mathcal{V}(x_{n+1}, \sigma_{n+1}) \leq \mathcal{V}(x_n, \sigma_{n+1})$

(b) So by substraction, we get
$V(x_{n+1}, \sigma_{n+1}) - \mathcal{V}(x_{n+1}, \sigma_n) \leq \mathcal{V}(x_n, \sigma_{n+1}) - \mathcal{V}(x_{n+1}, \sigma_n)$ so
$(\sigma_{n+1} - \sigma_n) \parallel C(x_{n+1}) \parallel^2 \leq (\sigma_{n+1} - \sigma_n) \parallel C(x_n) \parallel^2$

(c) is clear from (a) and (b)

(d) $\parallel C(x_n) \parallel^2 \leq \frac{2}{\sigma_n}[J(x^*) - J(x_n)] \leq \frac{2}{\sigma_n}[J(x^*) - J(x_1)] \to 0$

(e) From (c) we get $J(\overline{x}) \leq J(x^*)$ and $C(\overline{x}) = 0$.

Moreover, it can be shown that $\sigma_n C(x_n) \to \lambda^*$ (left to exercise)

Summary

- We consider constrained optimization problem with equality constraints
- We first detail Lagrange technique in the case of QP where this technique amounts to solve a linear system
- We generalize the method to ENLP.
- We introduce relaxation algorithms.
- We consider applications of QP to regularized inversion and to quadratic linear control problems.

For Further Reading I

- J.F.. Bonnans, J.C.. Gilbert, C. Lemaréchal, C.A. Sagastizabal.
  Numerical optimization, theoretical and practical aspects ,
  Part I, Springer (2009).
- R.Fletcher.
  Practical methods of optimization ,
  Wiley (1987).
- M.Samuelides.
  Optimisation numérique ,
  Polycopié SUPAERO (1998).