

# EN530.603 Applied Optimal Control

## Lecture 3: Constrained Optimization Basics

September 11, 2013

Lecturer: Marin Kobilarov

### 1 Equality Constraints

In optimal control we will encounter cost functions of two variables  $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  written as

$$L(x, u)$$

where  $x \in \mathbb{R}^n$  denotes the *state* and  $u \in \mathbb{R}^m$  denotes the *control inputs*. We are interested in minimizing this function subject to the *equality constraints*

$$f(x, u) = \begin{bmatrix} f_1(x, u) \\ \vdots \\ f_n(x, u) \end{bmatrix} = 0.$$

In order to establish optimality conditions we first differentiate the constraint  $f(x, u) = 0$  to get

$$df = f_x \cdot dx + f_u \cdot du = 0, \tag{1}$$

Then assuming  $f_x$  is a non-singular square matrix

$$dx = -f_x^{-1} f_u \cdot du,$$

i.e. this is how small changes in  $u$  (i.e.  $du$ ) must relate to small changes in  $x$  (i.e.  $dx$ ). Now we have that

$$dL = L_x \cdot dx + L_u \cdot du = (L_u - L_x f_x^{-1} f_u) du$$

which is interpreted as the gradient of  $L$  w.r.t.  $u$  at a point where  $f(x, u) = 0$  holds true. Recall that minimizing  $L$  with respect to  $u$  requires exactly that

$$L_u - L_x f_x^{-1} f_u = 0,$$

which is our first-order *necessary condition*. Notice that we assumed that the variables  $x$  and  $u$  are such that  $f_x$  is always nonsingular. This works well if the constraint  $f$  were linear but does not easily generalize.

#### 1.1 The Lagrangian multiplier approach

A more general approach is to “adjoin” the constraints to the cost using “multipliers”  $\lambda_1, \dots, \lambda_n$  to form a new function

$$H(x, u, \lambda) = L(x, u) + \sum_{i=1}^n \lambda_i f_i(x, u) \equiv L(x, u) + \lambda^T f(x, u),$$

where  $H$  is called the *Hamiltonian*. The idea is to transform the constraint optimization of  $L$  into an unconstrained minimization of the new function  $H$ .

We will now show that minimizing  $H$  is equivalent to solving the original problem. First note that the condition  $H_x = 0$  is equivalent to

$$L_x + \lambda^T f_x = 0 \quad \Rightarrow \quad \lambda^T = -L_x(f_x)^{-1},$$

so we would guess that this is the solution for  $\lambda$  as a function of  $x, u$  (and verify it later).

Keeping  $f(x, u) = 0$  fixed is equivalent to satisfying  $dx = -f_x^{-1} f_u \cdot du$  and we have

$$\begin{aligned} dL &= L_x dx + L_u du \\ &= (-L_x(f_x)^{-1} f_u + L_u) du \\ &= (L_u + \lambda^T f_u) du \\ &= H_u \cdot du \end{aligned} \tag{2}$$

Therefore, the condition  $dL = 0$  when  $f(x, u) = 0$  is equivalent to the *necessary optimality conditions* for  $H = L(x, u) + \lambda^T f(x, u)$ :

$$\partial_\lambda H = 0 \quad \Rightarrow \quad f(x, u) = 0, \tag{3}$$

$$\partial_x H = 0, \tag{4}$$

$$\partial_u H = 0 \tag{5}$$

which are  $2n + m$  equations for the  $2n + m$  unknowns  $x, u$ , and  $\lambda$ . Note that these equations are very general, e.g. they do not require finding coordinates  $x$  for which  $f_x$  must always be invertible.

### 1.1.1 Example

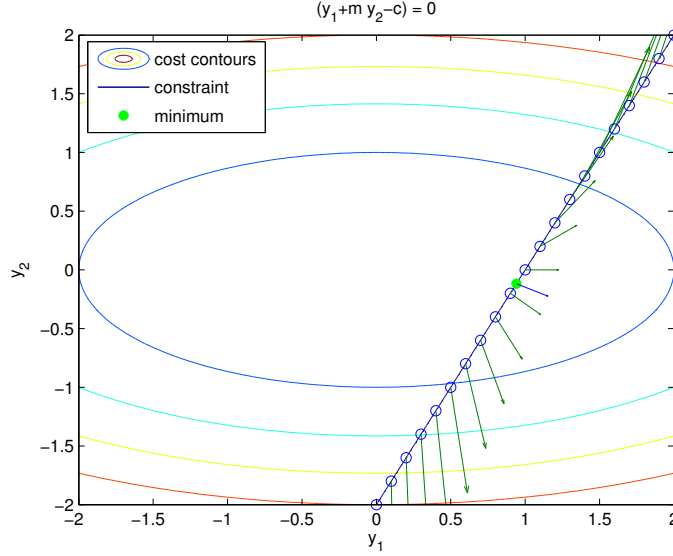
Consider  $L(x, u) = \frac{1}{2}qx^2 + \frac{1}{2}ru^2$  subject to  $f(x, u) = x + mu - c$ , where  $q > 0, r > 0, m, c$  are given constants. We have

$$H_x = qx + \lambda \quad \Rightarrow \quad \lambda = -qx \tag{6}$$

$$H_u = ru + m\lambda \quad \Rightarrow \quad u = -\frac{m}{r}\lambda = \frac{mq}{r}x \tag{7}$$

Substitute  $u$  into the constraint  $f(x, u) = 0$  we obtain

$$x + \frac{m^2q}{r}x - c = 0, \quad \Rightarrow \quad x = \frac{rc}{r + m^2q}$$



In order to determine the *sufficient* conditions we examine the second-order expansion of  $L(x, u)$ . This is most conveniently accomplished using  $L(x, u) = H(x, u, \lambda) - \lambda^T f(x, u)$ , i.e.

$$dL \approx (H_x, H_u) \begin{pmatrix} dx \\ du \end{pmatrix} + \frac{1}{2} (dx^T, du^T) \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{pmatrix} dx \\ du \end{pmatrix} - \lambda^T df$$

We can substitute the constraint

$$df = 0 \quad \Leftrightarrow \quad dx = -f_x^{-1} f_u du$$

as well as the necessary condition  $H_x = 0$  to obtain

$$dL \approx \frac{1}{2} du^T [-f_u^T (f_x^T)^{-1}, I] \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -f_x^{-1} f_u \\ I \end{bmatrix} du$$

The positive-definiteness of this quadratic form for all  $du \neq 0$  at an optimal solution  $u^*$  is a *sufficient condition* for a local optimum.

## 1.2 The general optimization setting

More generally, assume we want to minimize  $L(y)$ , for  $y \in \mathbb{R}^{n+m}$ , subject to  $n$  equalities

$$f(y) = \begin{bmatrix} f_1(y) \\ \vdots \\ f_n(y) \end{bmatrix} = 0$$

*Feasible* changes  $dy$  are tangent to  $f(y)$ , i.e. satisfy

$$f_y \cdot dy = 0,$$

which can also be equivalently written using gradient notation as:

$$\nabla f_i^T dy = 0, \quad \text{for all } i = 1, \dots, n.$$

We will employ geometric reasoning to obtain the optimality conditions. First, note that directions orthogonal to any feasible  $dy$  must be spanned by the gradients  $\{\nabla f_1, \dots, \nabla f_n\}$ . At an optimum  $y^*$  we must also have

$$\nabla L(y^*)^T dy = 0,$$

i.e.  $\nabla L$  is orthogonal to any feasible  $dy$  and must be spanned by gradients as well. This can be expressed as:

$$\nabla L(y^*) = - \sum_{i=1}^n \lambda_i^* \nabla f_i(y^*)$$

where the minus sign is by convention, and the scalars  $\lambda_i$  can be arbitrary. Therefore, we have

$$\nabla L(y^*) + \sum_{i=1}^n \lambda_i^* \nabla f_i(y^*) = 0 : \quad \text{:first-order necessary conditions}$$

along with

$$dy^T \left[ \nabla^2 L(y^*) + \sum_{i=1}^n \lambda_i^* \nabla^2 f_i(y^*) \right] dy \geq 0, \quad \text{:second-order necessary conditions}$$

then constitute the necessary conditions for optimality. Note that in the above  $dy$  is not arbitrary, i.e. we require that  $\nabla f_i(y^*)^T dy = 0$  for all  $i = 1, \dots, n$ .

Sufficient conditions for a strict local optimum are obtained by requiring the positive-definiteness of the quadratic form above.

Finally, note that the multipliers are related to the solution *sensitivity*. The relationship

$$\nabla L = - \sum_{i=1}^n \lambda_i \nabla f_i$$

signifies that the multipliers are, roughly speaking, the ratio of the change in cost to the change in constraint. In other words, the  $i$ -th multiplier  $\lambda_i$  determines how changes in the  $i$ -th constraint  $f_i$  relate to changes in the cost  $L$  as a result of perturbing the solution by  $dy$ .

## 2 Checking Sufficient Conditions in practice

A common algebraic way to check the sufficient conditions is through QR decomposition of the constraint gradient. In particular, we have that

$$\nabla f^T dy = 0 \quad \Rightarrow \quad dy \in \text{Null}(\nabla f^T) \tag{8}$$

and finding the null space can be accomplished using the QR decomposition of  $\nabla f$ , i.e.

$$\nabla f = Q \begin{bmatrix} R \\ 0 \end{bmatrix}.$$

where  $Q \in \mathbb{R}^{n \times n}$  is such that  $Q^T Q = I$  and  $R \in \mathbb{R}^{m \times m}$  is upper triangular. Therefore, any  $dy$  of the form

$$dy = Q \begin{bmatrix} 0 \\ du \end{bmatrix},$$

for some arbitrary  $du \in \mathbb{R}^{n-m}$  would satisfy (8). If we decompose  $Q$  according to

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix},$$

where  $Q_1 \in \mathbb{R}^{n \times m}$  and  $Q_2 \in \mathbb{R}^{n \times (n-m)}$ , then  $dy$  can be expressed using the last  $n - m$  columns in  $Q$  as

$$dy = Q_2 du,$$

which leads to the sufficient conditions

$$du^T Q_2^T \left[ \nabla^2 L(y^*) + \sum_{i=1}^m \lambda_i \nabla^2 f(y^*) \right] Q_2 du > 0,$$

for arbitrary  $du \neq 0$ , i.e. it reduces to checking that the quadratic matrix  $Q_2^T [\nabla^2 L(y^*) + \sum_{i=1}^m \lambda_i \nabla^2 f(y^*)] Q_2$  is positive definite.

### 3 Inequality Constraints

Inequality constraints are used to encode allowable regions in state and control space. A general class of problems with such constraints involve the minimization of

$$L(y)$$

subject to

$$f(y) \leq 0,$$

where  $f$  can be of any dimension. Let  $y^*$  be the unconstrained minimum of  $L(y)$ . If the constrained is not violated, i.e. if  $f(y^*) \leq 0$  then problem is solved. If we have that

$$f(y^*) > 0,$$

then we say that the constraints are *active* and must be enforced similar to equality constraints, i.e. using the Hamiltonian

$$H(y, \lambda) = L(y) + \lambda^T f(y),$$

with the main difference that the multipliers must be positive when the constraint is active, i.e.

$$\lambda = \begin{cases} \geq 0, & f(y) = 0, \\ = 0, & f(y) < 0. \end{cases}$$

The condition  $H_y = 0$  is equivalent to the relationship

$$\nabla L = - \sum_{i=1}^n \lambda_i \nabla f_i$$

which now has the geometric interpretation that the cost gradient must be spanned by the *negative* constraint gradients. In other words, *the gradient of  $L$  with respect to  $y$  at a minimum must be pointed in such a way that decrease of  $L$  can only come by violating the constraints.*

The *sufficient condition* for local minimum of  $L(y)$  with  $f(y) \leq 0$  includes the standard equality constraint conditions to which we add the condition that all  $\lambda > 0$ .

Note: when the constraint is active we let  $\lambda \geq 0$  rather than require  $\lambda > 0$  since the case  $\lambda = 0$  might also satisfy the necessary conditions. In fact, when  $\lambda = 0$  then  $\nabla L = 0$  which is more restrictive than only requiring the cost gradient to be spanned by constraint gradients.

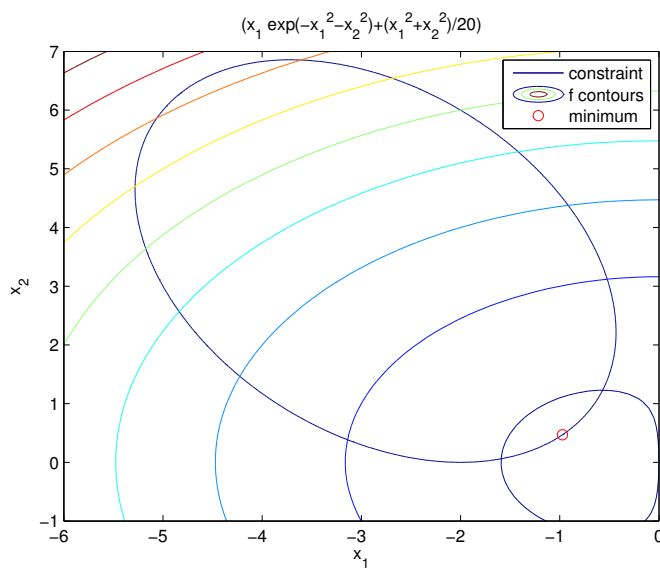
### 3.1 Example

Consider the function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$L(x) = x_1 \exp(-(x_1^2 + x_2^2)) + (x_1^2 + x_2^2)/20$$

subject to the inequality constraint

$$f(x) = x_1 x_2 / 2 + (x_1 + 2)^2 + (x_2 - 2)^2 / 2 - 2 \leq 0$$



See *lecture3.2.m*

### References