Constrained problems

Niclas Börlin

5DA001 Non-linear Optimization

The continuous optimization problem

- In its most general form, a continuous optimization problem may be written

$$\min_{x \in \Re^n} f(x) \qquad \text{subject to} \qquad \begin{array}{ll} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{array}$$

- The function $f(x) : \Re^n \rightarrow \Re$ is called the objective function and is assumed to be twice continuously differentiable.
- The vector x contains the variables to be estimated.
- The functions $c_i(x) : \Re^n \rightarrow \Re$ define constraints on the variables.
- The sets $\mathcal{E}$ and $\mathcal{I}$ are index sets for the equality and inequality constraints, respectively.
- A maximization problem is rewritten as

$$\max_x f(x) \equiv -\min_x -f(x).$$

Example

- Consider the problem

$$\begin{aligned} \min\ & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{subject to } & x_2 \geq x_1^2, \\ & x_1 < 2 - x_2. \end{aligned}$$

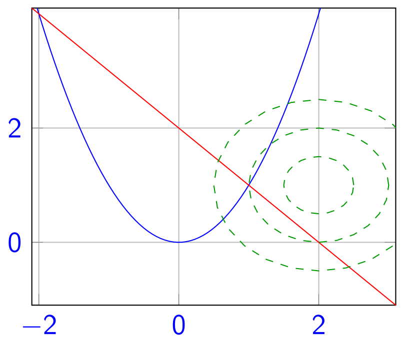
- We may rewrite this problem into general form as

$$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2,$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

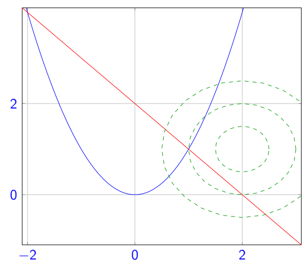
$$c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \begin{bmatrix} -x_1^2 + x_2 \\ -x_1 - x_2 + 2 \end{bmatrix},$$

$$\mathcal{I} = \{1, 2\}, \ \mathcal{E} = \{\}.$$



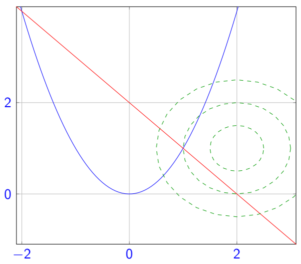
The parameter space

- The vector x will be interpreted as a point in $\Re^n$, the parameter space.
- Points that satisfies all constraints are called feasible and belong to the feasible set Ω which is a subset of $\Re^n$.
- At a feasible point x, an inequality constraint $c_i(x) \geq 0$ is said to be binding or active if $c_i(x) = 0$.
- If $c_i(x) > 0$, the constraint is nonbinding or inactive.
- Equality constraints are always active.



The parameter space

- The point x is said to be on the boundary of the constraint if $c_i(x) = 0$ and in the interior of the constraint if $c_i(x) > 0$.
- Equality constraints have no interior points.
- The set of active constraints at a given point is called the active set (of constraints).
- A feasible point with at least one active constraint belongs to the boundary of the feasible set.
- Feasible points with no active constraints are interior points to the feasible set.



Optimality conditions for constrained problems

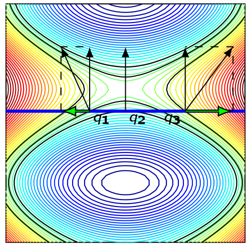
- A minimizer x^* to a minimization problem

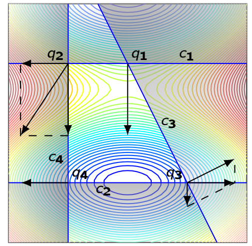
$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \\ & c_i(x) \geq 0, \quad i \in \mathcal{I} \end{aligned}$$

must satisfy

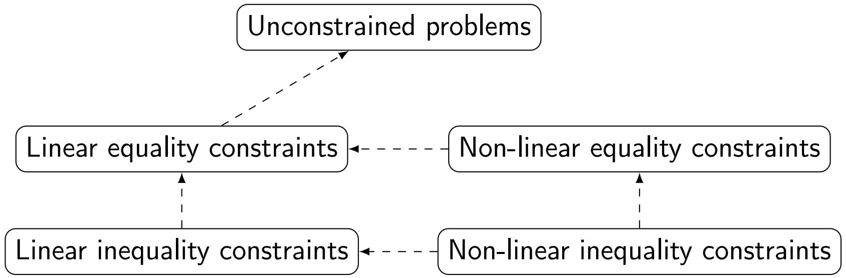
$$p^T \nabla f(x^*) \geq 0$$

for all feasible directions p.





Overview



Linear equality constraints

- Consider a problem with linear equality constraints, i.e.

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

where A is assumed to have full rank.
- The constrained problem may be rewritten to the unconstrained problem

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

where $\bar{x}$ is a feasible point and $Z \in \Re^{n \times r}$ is a basis for $\mathcal{N}(A)$.
- The function $\phi(v)$ is called the reduced function.

Necessary conditions for the reduced problem

- The necessary conditions for the reduced problem are

$$\begin{aligned} \nabla\phi(v) &= Z^T\nabla f(\bar{x}+Zv) = Z^T\nabla f(x) = 0, \\ \nabla^2\phi(v) &= Z^T\nabla^2 f(\bar{x}+Zv)Z = Z^T\nabla^2 f(x)Z \text{ pos. semidef.,} \end{aligned}$$

  where $x = \bar{x} + Zv$.
- The expression

$$Z^T\nabla f(x)$$

  is called the reduced gradient and

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

  the reduced Hessian.
- If the null space matrix Z is an orthogonal projection matrix, they are called projected gradient and Hessian, respectively.

Necessary conditions for the reduced problem
Cont'd

- The second order condition corresponds to

$$v^T Z^T \nabla^2 f(x^*) Z v \geq 0 \quad \forall v,$$

  which may be rewritten as

$$p^T \nabla^2 f(x^*) p \geq 0 \quad \forall p \in \mathcal{N}(A),$$

  where $p = Zv$.
- Thus, the Hessian in x^* has to be positive semidefinite on the null space of A.
- **NB:** This does not mean that the Hessian in x^* has to be positive semidefinite on the whole $\Re^n$.

Sufficient conditions for linear equality constraints

- If x^* satisfies

$$\boxed{\begin{aligned} & Ax^* = b, \\ & Z^T\nabla f(x^*) = 0, \text{ and} \\ & Z^T\nabla^2 f(x^*)Z \text{ is positive definite,} \end{aligned}}$$

  where Z is a basis matrix for $\mathcal{N}(A)$, then x^* is a strict local minimizer of f over $\{x : Ax = b\}$.

Example

- Consider the problem

$$\begin{aligned} & \min_x f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3, \\ & \text{s.t. } x_1 - x_2 + 2x_3 = 2, \end{aligned}$$

  with gradient and Hessian functions

$$\nabla f(x) = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 \\ -2x_3 + 4 \end{bmatrix} \text{ and } \nabla^2 f(x) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

- As null space matrix of the constraint matrix

$$A = [1, -1, 2]$$

  we may choose

$$Z = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Example

The reduced gradient

- In the feasible point

$$x = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix},$$

  the reduced gradient is

$$Z^T \nabla f(x) = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

  and x is not a local minimizer.
- In the feasible point

$$x^* = \begin{bmatrix} 2.5 \\ -1.5 \\ -1 \end{bmatrix}, \nabla f(x^*) = \begin{bmatrix} 3 \\ -3 \\ 6 \end{bmatrix}, \text{ and } Z^T \nabla f(x) = 0.$$

- Hence, x^* is potentially a local minimizer.

Example

The reduced Hessian

- The reduced Hessian in x^* is

$$Z^T \nabla^2 f(x^*) Z = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix}$$

  and is positive definite.
- Thus, the second order sufficient conditions are satisfied and x^* is a strict local minimizer of f.
- Notice that $\nabla^2 f(x)$ itself is not positive definite.

The Lagrange multipliers

- Let x^* be a minimizers and $Z \in \Re^{n \times r}$ a null space matrix for A.
- The gradient $\nabla f(x^*)$ may be expressed as the sum of its null space and range space components, i.e.

$$\nabla f(x^*) \quad = \quad Z v^* + A^T \lambda^*,$$

  where $v^* \in \Re^r$ and $\lambda^* \in \Re^m$.
- Together with the first order conditions we get

$$\begin{aligned} Z^T \nabla f(x^*) &= Z^T Z v^* + Z^T A^T \lambda^* = \\ &= Z^T Z v^* + (\underbrace{AZ}_{=0})^T \lambda^* \\ &= Z^T Z v^* = 0 \\ &\Downarrow \\ Z v^* &= 0. \end{aligned}$$

The Lagrange multipliers

- Thus, a minimizer satisfies

$$\nabla f(x^*) = A^T \lambda^*.$$

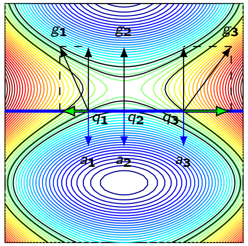
- In other words: In a local minimum, the gradient of the objective function is a linear combination of the gradients of the constraints.
- The coefficients in the vector λ^* are called Lagrange multipliers.
- The constraint and first order condition may be formulated in one system equation of $n + m$ equations and $n + m$ unknowns in x and λ:

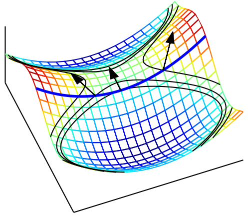
$$\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} \nabla f(x) \\ b \end{bmatrix}.$$

Example

$$\min_x \quad f(x) = x^2 + \sin^2 2y$$
$$\text{s.t.} \quad y = 0.6$$

$$A = \begin{bmatrix} 0 & -1 \end{bmatrix}, Z = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$
$$x^* = \begin{bmatrix} 0 \\ 0.6 \end{bmatrix}, \nabla f(x^*) = \begin{bmatrix} 0 \\ 1.35 \end{bmatrix},$$
$$\nabla^2 f(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & -5.9 \end{bmatrix},$$
$$Z^T \nabla^2 f(x^*) Z = \begin{bmatrix} 2 \end{bmatrix}.$$





The Lagrange multipliers and sensitivity

- The Lagrange multipliers may be used to estimate the sensitivity of the min value $f(x^*)$ with respect to the constraints.
- Assume we have found a solution x^* to

$$\min_x f(x)$$
$$\text{s.t. } Ax = b.$$

- Consider the perturbed constraints $Ax = b + \delta$.

The Lagrange multipliers and sensitivity

- If the perturbation δ is small enough, the solution $\bar{x}$ to the perturbed problem will be close to x^* and

$$\begin{aligned} f(\bar{x}) &\approx f(x^*) + (\bar{x} - x^*)^T \nabla f(x^*) = f(x^*) + (\bar{x} - x^*)^T A^T \lambda^* \\ &= f(x^*) + (A\bar{x} - Ax^*)^T \lambda^* = f(x^*) + (b + \delta - b)^T \lambda^* \\ &= f(x^*) + \delta^T \lambda^* \\ &= f(x^*) + \sum_{i=1}^m \delta_i \lambda_{*i}. \end{aligned}$$

- Thus, if element i of the right hand side of the constraint is modified by δ_i, the optimal objective value will change with about $\delta_i \lambda_{*i}$.
- For this reason, the Lagrange multipliers are sometimes called shadow prices or dual variables.

The Lagrangian function

- Define the Lagrangian function of x and λ as

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i (a_i^T x - b_i) = f(x) - \lambda^T (Ax - b).$$

- The gradients of the Lagrangian are

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

and

$$\nabla_\lambda \mathcal{L}(x, \lambda) = b - Ax.$$

- The first order condition on the Lagrangian

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

correspond to the first order condition on the constrained problem.
- A local minimizer to the constrained problem is a stationary point to the Lagrangian.

Optimality conditions for linear inequality constrained problems

- Consider a problem with linear inequality constraints, i.e.

$$\begin{aligned} &\min_x f(x) \\ &\text{s.t. } Ax \geq b, \end{aligned}$$

  where A is assumed to have full rank.
- The active constraints in a point x^* will determine if x^* is a minimizer.
- Our problem may thus be rewritten as

$$\begin{aligned} &\min_x f(x) \\ &\text{s.t. } \hat{A}x = \hat{b}, \end{aligned}$$

  where $\hat{A}$ and $\hat{b}$ contains the active constraints.

Optimality conditions for linear inequality constrained problems
Cont'd

- If Z is a null space matrix of $\hat{A}$, the first order condition becomes

$$Z^T \nabla f(x^*) = 0$$

  or

$$\nabla f(x^*) = \hat{A}^T \hat{\lambda}_*,$$

  where $\hat{\lambda}_* \geq 0$ has the Lagrange multipliers for the active constraints.
- The second order necessary condition is that $Z^T \nabla^2 f(x^*) Z$ must be positive semidefinite.

Complementary slackness

- If we define the Lagrange multiplier of an inactive constraint to be zero, we may describe the inequality conditions as

$$\lambda_{*i}(a_i^T x^* - b_i) = 0, \quad i = 1, \ldots, m.$$

- This condition is called complementary slackness and means that
  - either the constraint is active ($a_i^T x^* - b_i = 0$)
  - or the Lagrange multiplier is zero ($\lambda_{*i} = 0$).
- At least one of the two must be true.
- The case when both cannot be true at the same time is called strict complementarity.
- Without strict complementarity, a Lagrange multiplier may be zero even if the constraint is active.
- In such a case, that constraint is called degenerate.

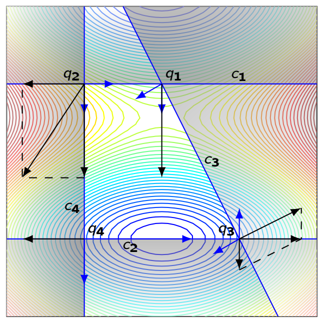
Example

For the problem

$$\begin{aligned} \min_x \quad & f(x) = x^2 + \sin^2 2y \\ \text{s.t.} \quad & \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq \begin{bmatrix} -1 \\ 0 \\ -1 \\ -1 \end{bmatrix} \end{aligned}$$

there are four corners, two of which are degenerate.

Point	Active constraints	(x,y)	∇f	λ
q_1	1,3	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -1.5 \end{bmatrix}$	$1.5, 0$
q_2	1,4	$\begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -1 \\ -1.5 \end{bmatrix}$	$1.5, -0.5$
q_3	2,3	$\begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$-0.5, -0.5$
q_4	2,4	$\begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 0 \end{bmatrix}$	$0, -0.5$



Necessary condition, linear inequalities

- In summary, the following conditions have to be satisfied for a point x^* to be a minimizer of a function f on the set $\{x : Ax \geq b\}$:

$$\boxed{\begin{aligned} & Ax^* \geq b, \\ & \nabla f(x^*) = A^T \lambda^* \Leftrightarrow Z^T \nabla f(x^*) = 0, \\ & \lambda^* \geq 0, \\ & \lambda^{*T}(Ax^* - b) = 0, \\ & Z^T \nabla^2 f(x^*) Z \text{ positive semidefinite,} \end{aligned}}$$

  for some vector λ^* of Lagrange multipliers and where Z is a null space matrix for the matrix $\hat{A}$ of the active constraints in x^*.

Sufficient conditions, linear inequalities I

- With strict complementarity we may extend to sufficient conditions in a straightforward manner:
- Assume x^* satisfies

$$\boxed{\begin{aligned} & Ax^* \geq b, \\ & \nabla f(x^*) = A^T \lambda^*, \\ & \lambda^* \geq 0, \\ & \text{we have strict complementarity,} \\ & Z^T \nabla^2 f(x^*) Z \text{ positive definite,} \end{aligned}}$$

- Then x^* is a strict local minimizer of f on the set $\{x : Ax \geq b\}$.

Why strict complementarity is needed

- The point x^* is also a strict local minimizer on the set $\{x : \hat{A}x = \hat{b}\}$, i.e. f increases in all directions p such that $\hat{A}p = 0$:
- Study a direction p such that $\hat{A}p \geq 0$, where some component of p is strictly positive, i.e. p points *into* the feasible set.
- Since

$$\nabla f(x^*) = A^T \lambda^* = \hat{A}^T \hat{\lambda}_*,$$

  then

$$p^T \nabla f(x^*) = p^T \hat{A}^T \hat{\lambda}_* = (\hat{A}p)^T \hat{\lambda}_*.$$

Why strict complementarity is needed
Cont'd

- With strict complementarity, we know that

$$(\hat{A}p)^T \hat{\lambda}_* > 0,$$

  i.e. p is an ascent direction and x^* must be a strict minimizer.
- Without strict complementarity,

$$(\hat{A}p)^T \hat{\lambda}_* = 0$$

  may be true for some p.
- This, we cannot tell anything about x^* with only first order information.
- However, if we drop the degenerate constraints, we may formulate sufficient conditions on the remaining constraints.

Sufficient conditions, linear inequalities II

- Let $\hat{A}_+$ contain the rows of $\hat{A}$ corresponding to the non-degenerate constraints in x^*.
- Let Z_+ be a null space matrix to $\hat{A}_+$.
- Assume x^* satisfies

$$\boxed{\begin{aligned} &Ax^* \geq b,\\ &\nabla f(x^*) = A^T\lambda^*,\\ &\lambda^* \geq 0,\\ &\lambda^{*T}(Ax^* - b) = 0,\\ &Z_+^T\nabla^2 f(x^*)Z_+ \text{ positive definite.} \end{aligned}}$$

- Then x^* is a strict local minimizer to the inequality constrained problem.

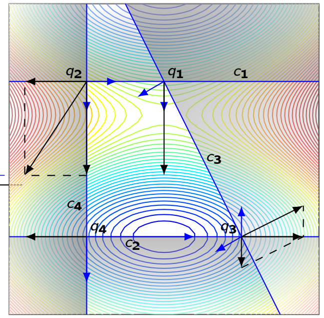
Example

For the problem

$$\begin{aligned} \min_x \quad & f(x) = x^2 + \sin^2 2y \\ \text{s.t.} \quad & \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -2 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq \begin{bmatrix} -1 \\ 0 \\ -1 \\ -1 \end{bmatrix} \end{aligned}$$

we have

	λ	$\nabla^2 f$	Z_+	$Z_+^T\nabla^2 f Z_+$
q_1	$\lambda_{1,3} = 1.5, 0$	$\begin{bmatrix} 2 & 0 \\ 0 & -5.2 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$[2]$
q_2	$\lambda_{1,2} = 1.5, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & -5.2 \end{bmatrix}$	$[]$	$[]$
q_3	$\lambda_{2,3} = -0.5, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$	$[]$	$[]$
q_4	$\lambda_{2,4} = 0, -0.5$	$\begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$[8]$



Optimality conditions for non-linear constraints

- Non-linear optimization problems with non-linear constraints are formulated as

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

for equality constraints, and

$$\begin{aligned} &\min_x f(x) \\ &\text{s.t. } c_i(x) \geq 0, i = 1, \ldots, m \end{aligned}$$

for inequality constraints.

- We will assume that the solution point x^* is regular, i.e. that the gradients of the active constraints in x^* $\{\nabla c_i(x^*) : c_i(x^*) = 0\}$ are linearly independent.

Optimality conditions for non-linear constraints

Cont'd

- The optimality conditions are expressed in terms of the Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) = f(x) - \lambda^T c(x),$$

where λ is a vector of Lagrange multipliers and c is a vector of constraint functions $\{c_i\}$.

Necessary conditions for equality constraints

- Let x^* be a local minimizer for f under the constraints $c(x) = 0$ and $Z(x^*)$ be a null space matrix for the Jacobian $\nabla c(x^*)^T$ of the constraints.
- If x^* is a regular point, then there exists a Lagrangian vector λ^* such that

$$\boxed{\begin{array}{l} \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \Leftrightarrow Z(x^*)^T \nabla f(x^*) = 0, \\ Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \text{ positive semi-definite.} \end{array}}$$

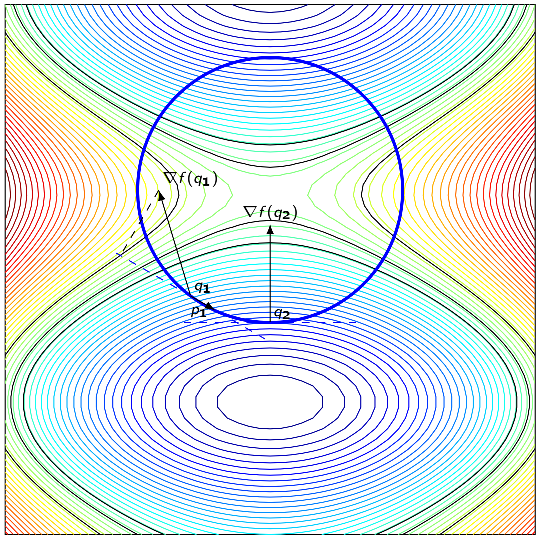
Sufficient conditions for equality constraints

- Let x^* be a point such that $c(x^*) = 0$ and $Z(x^*)$ is a basis for the null space of $\nabla c(x^*)^T$.
- Assume there exists a vector λ^* such that

$$\boxed{\begin{array}{l} \nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \\ Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \text{ is positive definite.} \end{array}}$$

- Then x^* is a strict local minimizer to f on the constraint set $\{x : c(x) = 0\}$.

Example



Linear constraints revisited

- For linear constraints $c(x) = Ax - b$, the Jacobian is $\nabla c(x)^T = A$ and the first order conditions

$$Z(x^*)^T \nabla f(x^*) = 0 \quad \Leftrightarrow \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \nabla c(x^*) \lambda = 0$$

becomes

$$Z^T \nabla f(x^*) = 0 \quad \Leftrightarrow \quad \nabla f(x^*) = A^T \lambda^*.$$

- The second order necessary (sufficient) conditions becomes that

$$\nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) = \nabla^2 f(x^*)$$

should be positive semi-definite (definite).

Necessary conditions for inequality constraints

- Let x^* be a local minimizer for f under the constraints $c(x) \geq 0$ and $Z(x^*)$ be a null space matrix for the Jacobian of the active constraints in x^*.
- If x^* is a regular point, then there exists a Lagrangian vector λ^* such that

$$\boxed{\begin{array}{l}\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \Leftrightarrow Z(x^*)^T \nabla f(x^*) = 0, \\ \lambda^* \geq 0, \\ {\lambda^*}^T c(x^*) = 0, \\ Z(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z(x^*) \text{ positive semi-definite.}\end{array}}$$

- The condition ${\lambda^*}^T c(x^*) = 0$ is the non-linear version of the complementary slackness condition.

Sufficient conditions for inequality constraints

- Let x^* be a points such that $c(x^*) \geq 0$.
- Assume there exists a vector λ^* such that

$$\boxed{\begin{array}{l}\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \\ \lambda^* \geq 0, \\ {\lambda^*}^T c(x^*) = 0, \\ Z_+(x^*)^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z_+(x^*) \text{ is positive definite,}\end{array}}$$

  where $Z_+(x^*)$ is a basis for the null space of the Jacobian of the non-degenerate constraints in x^*.
- Then x^* is a strict local minimizer to f on the constraint set $\{x : c(x) \geq 0\}$.
- The necessary and sufficient conditions for the non-linear inequality constraints are often called the KKT conditions (Karush-Kuhn-Tucker conditions).