Optimization

Màster de Fonaments de Ciència de Dades

Lecture VI. Penalty and barrier function methods for constrained optimization

Gerard Gómez

Table of contents

1. Penalty function methods

Penalty function methods. General idea

- Consider the following **constrained optimization problem**:

> Seek a minimum of a real-valued function f on a feasible set $X \subset \mathbb{R}^n$

- This is problem **can be transformed into an unconstrained optimization one** after some modification of the objective function f using penalty functions

- Define $\sigma : \mathbb{R}^n \to \mathbb{R}$ as

$$\sigma(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in X \\ +\infty & \text{if } \boldsymbol{x} \notin X \end{cases}$$

  The function σ is called the **infinite penalty function** (also the **"ideal penalty"**), for it imposes an (infinite) penalty on points lying outside the feasible set X

Penalty function methods. General idea

- Consider the unconstrained minimization of the **augmented objective function** F defined by

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) = \min_{\boldsymbol{x}\in\mathbb{R}^n} (f(\boldsymbol{x}) + \sigma(\boldsymbol{x}))$$

where f is assumed to be defined on $\mathbb{R}^n$

Then

$$\boxed{\boldsymbol{x}^* \text{ minimizes } F \text{ in } \mathbb{R}^n \Leftrightarrow \boldsymbol{x}^* \text{ minimizes } f \text{ in } X}$$

Penalty function methods. General idea

- In practice, the unconstrained optimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} (f(\boldsymbol{x}) + \sigma(\boldsymbol{x}))$$

  cannot be, in general, carried out because:
  - The **discontinuity of F** on the boundary of X
  - The **infinite values outside X**
- Replacing $+\infty$ by some large finite penalty will not simplify the problem, since the numerical difficulties would still remain
- The idea for solving these problems involves a **sequence of unconstrained minimization problems**
- In each problem of the sequence **a penalty parameter** is adjusted from one minimization to the next one
- The **sequence of unconstrained minima converges** to a feasible point of the constrained problem

Penalty functions method. Example

- Consider the problem

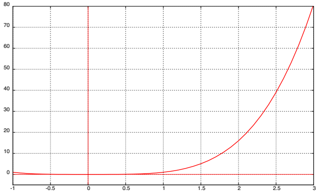
$$\begin{array}{ll} \min & f(x) = x^4 \\ \text{subject to} & g(x) = 1 - x \le 0 \quad (\Leftrightarrow \; x \ge 1) \end{array}$$

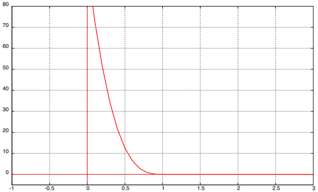
- Define a penalty function $\phi(t)$ by

$$\phi(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \lambda x^3 & \text{for} \quad x \ge 0 \end{cases}$$

where λ is some positive constant. Note that ϕ is twice differentiable, even through zero

- For minimization pourposes, the function $\phi(x)$ penalizes any number $x > 0$, and $\phi(1-x)$ any number $x \le 1$ (since $1 - x \ge 0 \;\Leftrightarrow\; x \le 1$)



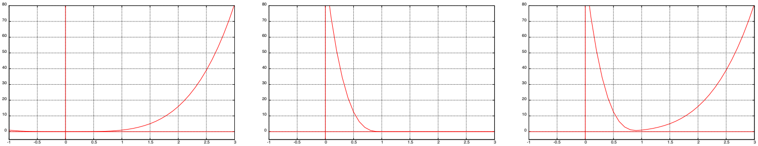


Graphs of $f(x)$ and $\phi(1-x)$ with $\lambda = 100$

Penalty functions method. Example

- Define a modified penalized objective function

$$\tilde{f}(x) = f(x) + \phi(1 - x)$$



$f(x)$ $\phi(1 - x)$ $\tilde{f}(x)$

- The function $\tilde{f}(x)$ is identical to f if $1 - x \leq 0$ ($\Leftrightarrow x \geq 1$), but rises sharply if $x < 1$

- The additional $\phi(1 - x)$ term penalizes an unconstrained optimization algorithm if $x < 1$

Penalty functions method. Example

- We can **approximately minimize** $f(x)$ subject to the constraint $x \geq 1$ **by running an unconstrained algorithm on the penalized objective function** $\tilde{f}(x)$
- BUT, running the golden section procedure on $\tilde{f}(x)$ (with $\lambda = 100$) we find that the minimum is at $x = 0.9012$. *Recall that the constraint is $x \geq 1$!!*
- The penalty approach didn't exactly solve the problem, but **it is reasonably close**
- The penalty term is also twice differentiable, so it should not cause any trouble in an optimization algorithm which relies on first or second derivatives

  The first and second derivatives of $\phi(x)$ are just

$$\phi'(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 3\lambda x^2 & \text{for} \quad x \geq 0 \end{cases} \qquad \phi''(t) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 6\lambda x & \text{for} \quad x \geq 0 \end{cases}$$

Penalty functions method. Example

- A reasonable procedure would be to **increase the constant λ, say by a factor of 10, and then re-run the unconstrained algorithm** on $\tilde{f}(x)$ using 0.9012 as the initial guess
- Increasing λ enforces the constrained more rigorously, while using the previous final iterate as an initial guess speeds up convergence (since we expect the minimum for the larger value of λ isn't that far from the minimum for the previous value of λ)
- In this case increasing λ to 10^4 moves the minimum to $x = 0.989$. *Recall that the constraint is $x \geq 1$!!*
- We could then increase λ and use $x = 0.989$ as an initial guess, and continue this process until we obtain a reasonable estimate of the minimizer

Penalty functions method. The general case

- In general we want to minimize a function $f(\boldsymbol{x})$ of n variables subject to both equality and inequality constraints of the form

$$\begin{aligned} g_i(\boldsymbol{x}) &\leq 0, \qquad i = 1, ..., p \\ h_j(\boldsymbol{x}) &= 0, \qquad j = 1, ..., m \end{aligned}$$

  Remark: If the constraint is $g_i(\boldsymbol{x}) \geq 0$ change $g_i(\boldsymbol{x})$ by $-g_i(\boldsymbol{x})$

- We will call $\phi(\lambda, t)$ for $\lambda \geq 0$ (parameter), $t \in \mathbb{R}$ (variable) a **penalty function** if

  1. $\phi(\lambda, t)$ is continuous
  2. $\phi(\lambda, t) \geq 0$ for all λ and t
  3. $\phi(\lambda, t) = 0$ for $t \leq 0$, and ϕ is strictly increasing for both $\lambda \geq 0$ and $t > 0$

  It's also desirable if ϕ has at least one continuous derivative in t (preferably two)

Penalty functions method. The general case

- A typical **example of a penalty function** would be

$$\phi(\lambda, t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ \lambda t^n & \text{for} \quad t \geq 0 \end{cases}$$

  where $n \geq 1$.

- This function has $n-1$ continuous derivatives in t, so taking $n = 3$ yields a C^2 penalty function.

Penalty functions method. The general case

- To minimize $f(x)$ subject to the above equality and inequality constraints, we define a modified objective function by

$$\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \phi(\alpha_i, g_i(\boldsymbol{x})) + \sum_{j=1}^{m} [\phi(\beta_j, h_j(\boldsymbol{x})) + \phi(\beta_j, -h_j(\boldsymbol{x}))]$$

  where the $\alpha_i > 0$ and $\beta_j > 0$ are constants that control how strongly the constraints will be enforced.

  In the modified objective function α_i and β_j play the role of the components of the vector parameter λ in $\phi(\lambda, t)$, and $g_i(\boldsymbol{x})$, $h_j(\boldsymbol{x})$ the role of the variable t

- The penalty functions ϕ in **the first sum** modify the original objective function so that **if any inequality constraint is violated ($g_i(\boldsymbol{x}) > 0$), a large penalty is invoked**; if all constraints are satisfied, no penalty
- Similarly **the second summation penalizes equality constraints** which are not satisfied, by penalizing both $h_j(\boldsymbol{x}) \leq 0$ and $h_j(\boldsymbol{x}) \geq 0$ (conditions that are equivalent to $h_j(\boldsymbol{x}) = 0$)

Penalty functions method. The general case

- We minimize the function $\tilde{f}(\boldsymbol{x})$ with no constraints, and count on the penalty terms to keep the solution near the feasible set although, in general, **no finite choice for the penalty parameters keeps the solution in the feasible set**.
- After having minimized $\tilde{f}(\boldsymbol{x})$ with an unconstrained method, for a given set of α_i and β_j, we may then increase the α_i and β_j and use the result of the last iterate as the initial guess for a new minimization
- Continue this process until we obtain a sufficiently accurate minimum

Penalty functions method

Example:

$$
\begin{array}{ll}
\min & f(\boldsymbol{x}) = x^2 + y^2 \\
\text{subject to} & g_1(x,y) = 6 - x - 2y \le 0 \\
& h_1(x,y) = 3 - x + y = 0
\end{array}
$$

- To start , we use the penalty function defined above with $\alpha_1 = 5$ and $\beta_1 = 5$

$$
\phi(\lambda, t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ \lambda t^n & \text{for} \quad t \ge 0 \end{cases}
$$

- The modified objective function is

$$
\begin{aligned}
\tilde{f}(x,y) &= f(x,y) + \phi(5, g_1(x,y)) + \phi(5, h_1(x,y)) + \phi(5, -h_1(x,y)) \\
&= x^2 + y^2 + \phi(5, 6 - x - 2y) + \phi(5, 3 - x + y) + \phi(5, -3 + x - y)
\end{aligned}
$$

Penalty functions method. Example

- Running the golden section algorithm on this, we get that the minimum occurs at
$$x = 3.506, \quad y = 1.001$$

  Note that the inequality and the equality constraints are violated
$$6 - x - 2y = 0.449 > 0, \quad 3 - x + y = 0.494$$

- To increase the accuracy we must increase the penalty parameters
- At each step, we use the final estimate from the previous penalty parameters as the initial guess for the larger parameters
- With $\alpha_1 = \beta_1 = 50$ we obtain $x = 3.836$, $y = 1.008$
- Increasing $\alpha_1 = \beta_1 = 500$ we obtain $x = 3.947$, $y = 1.003$

  Recal that the constraints are: $x + 2y \geq 6$ and $x - y = 3$, and $3.947 + 2 \cdot 1.003 = 5.953$, $3.947 - 1.003 = 2.944$
- The **solution of the problem is** $x = 4$, $y = 1$

Penalty functions method. Increasing the penalty parameters

- **Increasing the penalty parameters**
  - **Improves the accuracy** of the final answer
  - **Lows down** the unconstrained algorithm's **convergence**
- If we **increase the values of the parameters**, then $\tilde{f}(x)$ will have a very **large gradient** and the algorithm will spend a lot of time hunting for an accurate minimum
- Under appropriate assumptions, we will prove that **as the penalty parameters are increased without bound, any convergent subsequence of solutions to the unconstrained penalized problems must converge to a solution of the original constrained problem**

Pros and cons of penalty functions

Pros:

- The obvious advantage to the penalty function approach is that we obtain **a "hands-off" method for converting constrained problems of any type into unconstrained problems**
- We **don't have to worry about finding an initial feasible point** (sometimes is a problem)
- Many **constraints** in the real world are "**soft**", in the sense that they need not be satisfied precisely. The penalty function approach is well-suited to these type of problems

Pros and cons of penalty functions

Cons:

- The drawback to penalty function methods is that **the solution to the unconstrained penalized problem will not be an exact solution to the original problem** (except in the limit, as mentioned above)
- In some cases, penalty methods **can't be applied** because **the objective function is actually undefined outside the feasible set**
- As we **increase the penalty parameters** to more strictly enforce the constraints, the **unconstrained formulation becomes very ill-conditioned**, with large gradients and abrupt function changes

There are more efficient methods for approaching constrained optimization problems; we will see some of them next

2. Barrier function methods

Barrier function methods

- **Barrier function methods** are closely related to penalty function methods and, in fact, might as well be considered a type of penalty function method
- These methods are generally applicable **only to inequality constrained** optimization problems
- Barrier methods have the advantage that they **always maintain feasible iterates**, unlike the penalty methods above

Barrier function methods

Consider the nonlinear inequality-constrained problem

$$\begin{array}{lll} \min & f(\boldsymbol{x}), & \boldsymbol{x} \in \mathbb{R}^n \\ \text{subject to} & g_i(\boldsymbol{x}) \geq 0, & i = 1, \ldots, p \end{array}$$

Remark: If the constraint is $g_i(\boldsymbol{x}) \leq 0$ change $g_i(\boldsymbol{x})$ by $-g_i(\boldsymbol{x})$

- Barrier methods are strictly feasible methods. This is done by creating a barrier that keeps the iterates in the interior of the feasible region
- The methods use a barrier term that approaches the infinite penalty function $\sigma(\boldsymbol{x})$ already defined
- We will assume that the feasible set has a nonempty interior; that is, there exists some point $\boldsymbol{x}_0$ such that $g_i(\boldsymbol{x}_0) > 0$ for $i = 1, \ldots, p$
- We will also assume that it is possible to reach any boundary point of X by approaching it from the interior

The barrier function

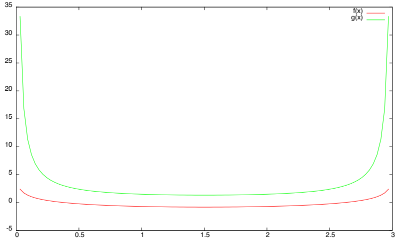
- Let $\phi(\boldsymbol{x})$ be a function that is continuous on the interior of the feasible set $\mathring{X}$, and that becomes unbounded as the boundary of the set ∂X is approached from its interior. This is

$$\text{if} \quad \boldsymbol{x} \to \partial X, \ \text{with} \quad \boldsymbol{x} \in \mathring{X}, \quad \text{then} \quad \phi(\boldsymbol{x}) \to \infty$$

- Two examples of such a function are the:

$$\text{logarithmic function} \quad \phi(\boldsymbol{x}) = -\sum_{i=1}^{p} \log(g_i(\boldsymbol{x})), \quad [-\log(x) - log(3-x) = f(x)]$$

$$\text{inverse function} \quad \phi(\boldsymbol{x}) = \sum_{i=1}^{p} \frac{1}{g_i(\boldsymbol{x})}, \quad \left[\frac{1}{x} + \frac{1}{3-x} = g(x)\right]$$



The barrier function

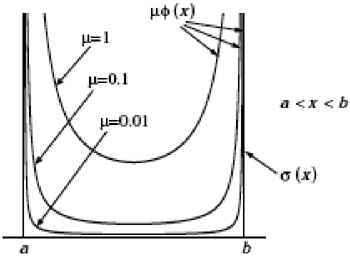
If, for instance, the constraint is $0 < x < 3$, this is

$$g_1(x) = x > 0 \quad \text{and} \quad g_2(x) = 3 - x > 0$$

then we can use as barrier functions

$$\phi(x) = -\log(x) - log(3 - x) \quad \text{and} \quad \phi(x) = \frac{1}{x} + \frac{1}{3 - x}$$

- Let μ be a positive scalar, then $\mu\phi(\boldsymbol{x})$ will "approach" $\sigma(\boldsymbol{x})$ as $\mu \to 0$



Effect of the barrier term $\mu\phi(\boldsymbol{x})$, as $\mu \to 0$, for $X = [a, b]$

Barrier function methods

- Adding a barrier term of the form $\mu\phi(\boldsymbol{x})$ to the objective function, we get

$$\tilde{f}_\mu(\boldsymbol{x}) = f(\boldsymbol{x}) + \mu\phi(\boldsymbol{x})$$

where $\mu > 0$ is referred to as the barrier parameter

- For instance, using the logarithmic barrier function we get

$$\tilde{f}_\mu(\boldsymbol{x}) = f(\boldsymbol{x}) - \mu \sum_{i=1}^{p} \log(g_i(\boldsymbol{x}))$$

- Barrier methods solve a sequence of unconstrained minimization problems of the form

$$\min_{\boldsymbol{x}} \ \tilde{f}_{\mu_k}(\boldsymbol{x}), \ k = 0, 1, 2, \ldots$$

for a sequence $\{\mu_k\}$ of positive barrier parameters that decrease monotonically to zero

Barrier function methods. General procedure

A barrier method works in a similar way to the penalty methods

1. We start with some **positive** μ **and feasible point** $\boldsymbol{x}_0$
2. **Minimize** $\tilde{f}$ using an **unconstrained algorithm**, **using**, if necessary, the **first-order necessary conditions for optimality**
3. Next, **decrease the value of the** μ and **re-optimize**, using the final iterate as an initial guess for the newly decreased μ
4. Continue until an acceptable minimum is found

Barrier function methods. Examples

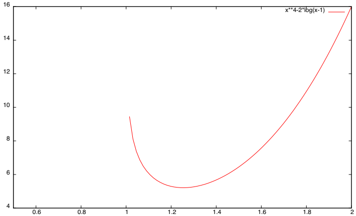
Example 1: Consider the objective function $f(t) = t^4$ and the constraint $t \geq 1$, this is, $g(t) = 1 - t \leq 0$.

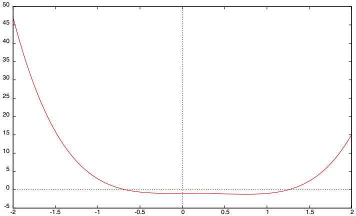
Taking $\mu = 2$, the penalized objective function is

$$\tilde{f}(t) = t^4 - 2\ln(t-1)$$

According to the necessary condition for optimality

$$\tilde{f}'(t) = 0 \quad \Rightarrow \quad 4t^3 - \frac{2}{t-1} = 0 \quad \Rightarrow \quad 2t^4 - 2t^3 - 1 = 0$$





$\tilde{f}(t) = t^4 - 2\ln(t-1)$ $\qquad$ $\tilde{f}'(t) = 2t^4 - 2t^3 - 1$

Barrier function methods. Examples (cont.)

Example 2: Let

$$f(x,y) = x^2 + y^2$$

We want to minimize f subject to $6 - x - 2y \leq 0$

- If we take $\mu = 5$ in the definition of $\tilde{f}$, so

$$\tilde{f}(x,y) = x^2 + y^2 - 5\ln(x + 2y - 6)$$

 and start with feasible point $(5,5)$, we obtain a minimum at $(x,y) = (1.53, 3.05)$

- Decreasing μ to 0.5, so

$$\tilde{f}(x,y) = x^2 + y^2 - 0.5\ln(x + 2y - 6)$$

 and starting at $(x,y) = (1.53, 3.05)$, gives a minimum at $(x,y) = (1.24, 2.48)$

- Decreasing μ to 0.05 and starting at $(x,y) = (1.24, 2.48)$, gives a minimum at $(x,y) = (1.204, 2.408)$

- The **solution of the problem** is $(x^*, y^*) = (1.2, 2.4)$

Barrier function methods. Examples (cont.)

Example 3: Solve

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = x - 2y \\ \text{subject to} & 1 + x - y^2 \geq 0 \\ & y \geq 0 \end{array}$$

- Then the logarithmic barrier function gives the unconstrained problem

$$\min_{\boldsymbol{x}} \tilde{f}_\mu(x, y) = x - 2y - \mu \log(1 + x - y^2) - \mu \log y$$

  for a sequence of decreasing barrier parameters
- For a specific parameter μ, the first-order necessary conditions for optimality are

$$\frac{\partial \tilde{f}_\mu(x, y)}{\partial x} = 1 - \frac{\mu}{1 + x - y^2} = 0, \quad \frac{\partial \tilde{f}_\mu(x, y)}{\partial y} = -2 + \frac{2\mu y}{1 + x - y^2} - \frac{\mu}{y} = 0$$

- Isolating x from the first equation ($x = y^2 - 1 + \mu$), from the second one we get

$$y^2 - y - \frac{\mu}{2} = 0$$

Barrier function methods. Examples (cont.)

- We can solve $y^2 - y - \frac{\mu}{2} = 0$ and, discarding the negative root and using $x = y^2 - 1 + \mu$, we get

$$y(\mu) = \frac{1 + \sqrt{1 + 2\mu}}{2}, \quad x(\mu) = \frac{\sqrt{1 + 2\mu} + 3\mu - 1}{2}$$

- As μ approaches zero, we obtain

$$\lim_{\mu \to 0_+} x(\mu) = \frac{\sqrt{1 + 2 \cdot 0} + 3 \cdot 0 - 1}{2} = 0, \quad \lim_{\mu \to 0_+} y(\mu) = \frac{1 + \sqrt{1 + 2 \cdot 0}}{2} = 1$$

- It is easy to verify that $\boldsymbol{x}^* = (0, 1)^T$ is indeed the solution to the problem

Remarks:

- One **issue** in using a barrier method is that of **finding an initial feasible point** which is in the interior of the feasible region. In many cases such a point will be obvious from considerations specific to the problem. If not, it can be rather difficult to find such a point.

- One idea **to find an initial point** would be to **use penalty functions, but on constraints** $g_i(\boldsymbol{x}) \leq -\delta < 0$ with $f = 0$. If a solution to this problem can be found with $\tilde{f}(\boldsymbol{a}) = 0$ then $\boldsymbol{a}$ is a feasible point which is in the interior of the feasible region defined by $g_i(\boldsymbol{x}) < 0$

Barrier function methods

The above example illustrates a number of typical features of barrier methods

1. It is possible to prove convergence for barrier methods under mild conditions of minimizers $\boldsymbol{x}(\mu)$ to the optimal solution $\boldsymbol{x}^*$

2. The sequence of barrier minimizers defines a differentiable curve $\boldsymbol{x}(\mu)$ known as the **barrier trajectory**

3. The barrier trajectory exists when the logarithmic or inverse barrier functions are used, provided that $\boldsymbol{x}^*$ is a regular point of the constraints[1] that satisfies the second-order sufficiency conditions, as well as the strict complementarity conditions (that will not be defined here)

4. The barrier trajectory can be used to develop techniques that accelerate the convergence of a barrier method

---

[1] In the case of equality constraints, $\boldsymbol{x}^*$ is a regular point if the gradients of the constraints are linearly independent. In the case of inequality constraints this means that the gradients of the active constraints at $\boldsymbol{x}^*$ $(g_i(\boldsymbol{x}^*) = 0)$ are linearly independent

Barrier function methods

Another **important feature** if the following

- Consider a point $\boldsymbol{x}(\mu)$ that is a minimizer of the logarithmic barrier function (similar for the invers barrier function) for a specific barrier parameter $\mu > 0$
- Setting the gradient of the barrier function to zero, we obtain

$$\nabla f(x) - \mu \sum_{i=1}^{p} \frac{\nabla g_i(\boldsymbol{x})}{g_i(\boldsymbol{x})} = 0, \quad \Rightarrow \quad \nabla f(x) - \sum_{i=1}^{p} \lambda_i(\mu) \nabla g_i(\boldsymbol{x}) = 0$$

  where

$$\lambda_i(\mu) = \frac{\mu}{g_i(\boldsymbol{x})}$$

- Therefore, we have a feasible point $\boldsymbol{x}(\mu)$ and a vector $\lambda(\mu)$ that satisfy the following relations

$$\begin{array}{rcll} \nabla f(x(\mu)) - \sum_{i=1}^{p} \lambda_i(\mu) \nabla g_i(\boldsymbol{x}(\mu)) & = & 0 & \\ \lambda_i(\mu) g_i(\boldsymbol{x}(\mu)) & = & \mu, & i = 1, \dots, p \\ \lambda_i(\mu) & \geq & 0, & i = 1, \dots, p \end{array}$$

Barrier function methods

- The above three relations **resemble the first-order necessary conditions for optimality of the constrained problem** (see Lecture 5)
- The only difference is that the condition

$$\lambda_i(\mu)g_i(\boldsymbol{x}(\mu)) = 0, \quad i = 1, ..., p$$

  is now replace by

$$\lambda_i(\mu)g_i(\boldsymbol{x}(\mu)) = \mu, \quad i = 1, ..., p$$

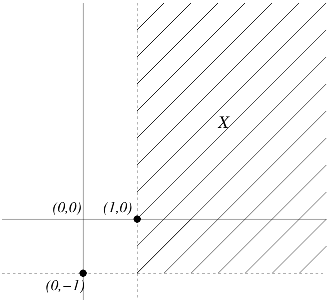
- Thus, $\boldsymbol{\lambda}(\mu)$ can be viewed as an estimate of the Lagrange multiplier $\boldsymbol{\lambda}^*$ at the optimal point $\boldsymbol{x}^*$
- Indeed, **if $\boldsymbol{x}^*$ is a regular point of the constraints, then as $\boldsymbol{x}(\mu)$ converges to $\boldsymbol{x}^*$, and $\boldsymbol{\lambda}(\mu)$ converges to $\boldsymbol{\lambda}^*$**
- The above results show that the points on the barrier trajectory, together with their associated Lagrange multiplier estimates, are the solutions of a perturbation of the first-order optimality conditions

Barrier function methods. Example

Example: Solve

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = x^2 + y^2 \\ \text{subject to} & g_1(x,y) = x - 1 \geq 0 \\ & g_2(x,y) = y + 1 \geq 0 \end{array}$$

- The solution to this problem is $\boldsymbol{x}^* = (1,0)^T$



Barrier function methods. Example (cont.)

- The first constraint $g_1(x,y) = x - 1 \geq 0$ is active at $\boldsymbol{x}^* = (1,0)^T$ ($g_1(\boldsymbol{x}^*) = 0$), and the corresponding Lagrange multiplier is $\lambda_1^* = 2$.

  The second constraint $g_2(x,y) = y + 1 \geq 0$ is inactive at $\boldsymbol{x}^* = (1,0)^T$ ($g_2(\boldsymbol{x}^*) > 0$), hence its Lagrange multiplier is $\lambda_2^* = 0$

  This is: $I(\boldsymbol{x}^*) = \{1\}$ (see Lecture 5), and $\boldsymbol{\lambda}^* = (2,0)^T$

- If the problem is solved using a logarithmic barrier function, the procedure solves the unconstrained minimization problem

$$\min \quad \tilde{f}_\mu(\boldsymbol{x}) = x^2 + y^2 - \mu \log(x-1) - \mu \log(y+1)$$

  for a decreasing sequence of barrier parameters μ that converge to zero

Barrier function methods. Example (cont.)

- The first-order necessary conditions for optimality are

$$\frac{\partial \tilde{f}_\mu(\boldsymbol{x})}{\partial x} = 2x - \frac{\mu}{x-1} = 0, \quad \frac{\partial \tilde{f}_\mu(\boldsymbol{x})}{\partial y} = 2y - \frac{\mu}{y+1} = 0$$

  yielding the unconstrained minimizers

$$x(\mu) = \frac{1+\sqrt{1+2\mu}}{2}, \quad y(\mu) = \frac{-1+\sqrt{1+2\mu}}{2}$$

- The Lagrange multiplier estimates at this point are

$$\lambda_1(\mu) = \frac{\mu}{g_1(\boldsymbol{x})} = \frac{\mu}{x-1} = \frac{2\mu}{\sqrt{1+2\mu}-1} = \sqrt{1+2\mu}+1$$

$$\lambda_2(\mu) = \frac{\mu}{g_2(\boldsymbol{x})} = \frac{\mu}{y+1} = \frac{2\mu}{\sqrt{1+2\mu}+1} = \sqrt{1+2\mu}-1$$

- When μ approaches zero, we obtain

$$\lim_{\mu \to 0} x(\mu) = 1, \quad \lim_{\mu \to 0} y(\mu) = 0$$

$$\lim_{\mu \to 0} \lambda_1(\mu) = 2, \quad \lim_{\mu \to 0} \lambda_2(\mu) = 0$$

- Thus

$$\boldsymbol{x}(\mu) \to \boldsymbol{x}^*, \quad \boldsymbol{\lambda}(\mu) \to \boldsymbol{\lambda}^*$$

Barrier function methods. Pros and cons

- A desirable property shared by both the logarithmic barrier function and the inverse barrier function is that the barrier function is convex if the constrained problem is a **convex optimization problem** defined in terms of a **convex objective function** and **concave constraint functions** (see Lecture 7)

- Barrier methods also have potential difficulties

- The unconstrained problems that appear usin barrier function methods become increasingly difficult to solve as the barrier parameter decreases

- The reason is that (with the exception of some special cases) the **condition number** of the Hessian matrix of the barrier function at its minimum point becomes increasingly large, tending to infinity as the barrier parameter tends to zero

The condition number

Assume we want to solve the linear system

$$A\boldsymbol{x} = \boldsymbol{b}$$

and both A and b are known with uncertainties δA and $\delta\boldsymbol{b}$ so, in fact we solve

$$(A + \delta A)(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b}$$

The condition number of the matrix A is defined as

$$\kappa = \|A^{-1}\| \, \|A\|$$

and is an amplifying factor of the relative errors

- If $\delta A = 0$, then
$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \kappa \frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$$
- If $\delta b = 0$, then
$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \frac{\kappa}{1 - \kappa\|\delta A\|/\|A\|} \frac{\|\delta A\|}{\|A\|}$$
- In general
$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le \frac{\kappa}{1 - \kappa\|\delta A\|/\|A\|} \left(\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\delta A\|}{\|A\|} \right)$$

Barrier function methods. Example

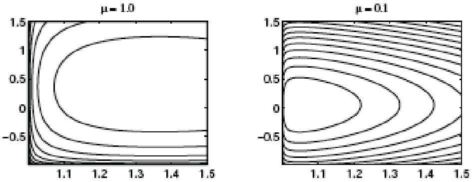
Example: Consider again the problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = x^2 + y^2 \\ \text{subject to} & g_1(x,y) = x - 1 \geq 0 \\ & g_2(x,y) = y + 1 \geq 0 \end{array}$$

with

$$\tilde{f}_\mu(\boldsymbol{x}) = x^2 + y^2 - \mu \log(x-1) - \mu \log(y+1)$$

Recall that the solution to this problem is $\boldsymbol{x}^* = (1,0)^T$



Contours $\tilde{f}_\mu(\boldsymbol{x}) = $ *ctant. for* $\mu = 1$ *and* $\mu = 0.1$

We see that for the smaller barrier parameter μ, the contours of the barrier function are almost parallel to the line $x = 1$, in fact they are almost parallel to the null space of the gradient of the active constraint g_1 at $\boldsymbol{x}^*$:

$$\nabla g_1(\boldsymbol{x}^*) = (1\ 0)^T,\ \nabla g_1(\boldsymbol{x}^*)(z_1\ z_2)^T = 0 \Rightarrow z_1 = 0$$

Barrier function methods. Example (cont.)

If

$$\tilde{f}_\mu(\boldsymbol{x}) = x^2 + y^2 - \mu \log(x-1) - \mu \log(y+1)$$

then

$$\nabla^2_{\boldsymbol{xx}} \tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} 2 + \frac{\mu}{(x-1)^2} & 0 \\ 0 & 2 + \frac{\mu}{(y+1)^2} \end{pmatrix}$$

Suppose now that $\boldsymbol{x}(\mu) = (x, y)^T$ is a minimizer of the barrier function for some value of μ

Recall that, for this problem, the Lagrange multipliers (see page 37) are

$$\lambda_1(\mu) = \frac{\mu}{g_1(\boldsymbol{x})} = \frac{\mu}{x-1} = \sqrt{1+2\mu}+1, \quad \lambda_2(\mu) = \frac{\mu}{g_2(\boldsymbol{x})} = \frac{\mu}{y+1} = \sqrt{1+2\mu}-1$$

and if μ is small

$$\lambda_1(\mu) \approx 2, \quad \lambda_2(\mu) \approx 0$$

Barrier function methods. Example (cont.)

Therefore

$$\nabla^2_{xx}\tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} 2 + \frac{\lambda_1^2(\mu)}{\mu} & 0 \\ 0 & 2 + \frac{\lambda_2^2(\mu)}{\mu} \end{pmatrix} \approx \begin{pmatrix} 2 + \frac{4}{\mu} & 0 \\ 0 & 2 \end{pmatrix}$$

The condition number κ of the Hessian matrix is approximately equal to

$$\frac{2 + 4/\mu}{2} = 1 + \frac{2}{\mu} = O\left(\frac{1}{\mu}\right)$$

(exercise) hence the matrix is ill conditioned

The **ill-conditioning of the Hessian matrix of the barrier function has several consequences**

- It rules out the use of an unconstrained method whose convergence rate depends on the condition number of the Hessian matrix at the solution
- Newton-type methods are sensitive to the ill-conditioning of the Hessian matrix and the numerical errors can result in a poor search direction

3. Penalty function methods for equality constrained problems

Penalty function methods for equally constrained problems

- Consider the problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = 0 \end{array}$$

  where $\boldsymbol{h}(\boldsymbol{x})$ is an m-dimensional vector whose i-th component is $h_i(\boldsymbol{x})$. We assume that all functions are twice continuously differentiable

- In general, the penalty function for constraint violation will be a continuous function $\phi(\boldsymbol{x})$ with the property that

$$\begin{array}{ll} \phi(\boldsymbol{x}) = 0 & \text{if } \boldsymbol{x} \text{ is feasible} \\ \phi(\boldsymbol{x}) > 0 & \text{otherwise} \end{array}$$

- The best-known such penalty is the quadratic-loss function

$$\phi(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{m} h_i^2(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{h}(\boldsymbol{x})$$

Penalty function methods for equally constrained problems

- The weight of the penalty is a positive penalty parameter ρ

  As ρ increases, the function $\rho\phi(\boldsymbol{x})$ approaches the **infinite penalty function ("ideal penalty")**:

$$\sigma(\boldsymbol{x}) = 0 \quad \text{if} \quad \boldsymbol{x} \in X \quad \text{and} \quad \sigma(\boldsymbol{x}) = \infty \quad \text{otherwise}$$

- By adding the term $\rho\phi(\boldsymbol{x})$ to $f(\boldsymbol{x})$ we obtain the penalty function

$$\tilde{f}_{\rho_k}(\boldsymbol{x}) = f(\boldsymbol{x}) + \rho\phi(\boldsymbol{x})$$

- The **penalty method** consists of solving a **sequence of unconstrained minimization problems** of the form

$$\min \quad \tilde{f}_{\rho}(\boldsymbol{x})$$

  for an increasing sequence $\{\rho_k\}$ of positive values tending to infinity

Penalty function methods for equally constrained problems

- As we have already seen, in general, the minimizers of the penalty function violate the constraints $\boldsymbol{h}(\boldsymbol{x})$. The growing penalty gradually forces these minimizers towards the feasible region
- Penalty methods share many of the properties of barrier methods
- Under mild conditions, it is possible to guarantee convergence
- Also, under appropriate conditions, the sequence of penalty function minimizers defines a continuous trajectory

Penalty function methods for equally constrained problems

- Consider, for example, the quadratic-loss penalty function

$$\tilde{f}_\rho(\boldsymbol{x}) = f(\boldsymbol{x}) + \frac{1}{2}\rho \sum_{i=1}^{m} h_i^2(\boldsymbol{x})$$

- The minimizer $\boldsymbol{x}(\rho)$ of $\tilde{f}_\rho(\boldsymbol{x})$ satisfies

$$\nabla \tilde{f}_\rho(\boldsymbol{x}(\rho)) = \nabla f(\boldsymbol{x}(\rho)) + \rho \sum_{i=1}^{m} \nabla h_i(\boldsymbol{x}(\rho))\, h_i(\boldsymbol{x}(\rho)) = 0$$

- Defining the Lagrange multiplier estimate

$$\lambda_i(\rho) = -\rho h_i(\boldsymbol{x}(\rho))$$

  we can write

$$\nabla f(\boldsymbol{x}(\rho)) - \sum_{i=1}^{m} \lambda_i(\rho)\, \nabla h_i(\boldsymbol{x}(\rho)) = 0$$

- If $\boldsymbol{x}(\rho)$ converges to a solution $\boldsymbol{x}^*$ that is a regular point of the constraints, then $\boldsymbol{\lambda}(\rho)$ converges to the Lagrange multiplier $\boldsymbol{\lambda}^*$ associated with $\boldsymbol{x}^*$

Penalty function methods for equally constrained problems. Example

Example: Solve using the quadratic-loss penalty function

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = -xy \\ \text{subject to} & h(x,y) = x + 2y - 4 = 0 \end{array}$$

- We must solve the sequence of unconstrained minimization problems

  $$\min \quad \tilde{f}_\rho(\boldsymbol{x}) = -xy + \frac{1}{2}\rho(x + 2y - 4)^2$$

  for increasing values of the penalty parameter ρ

- The necessary conditions for optimality for the unconstrained problem are

  $$\begin{aligned} \frac{\partial \tilde{f}_\rho(\boldsymbol{x})}{\partial x} = -y + \rho(x + 2y - 4) &= 0 \\ \frac{\partial \tilde{f}_\rho(\boldsymbol{x})}{\partial y} = -x + 2\rho(x + 2y - 4) &= 0 \end{aligned}$$

- For $\rho > 1/4$ (the unconstrained problem has no minimum if $\rho \leq 1/4$) this yields the solution

  $$x(\rho) = \frac{8\rho}{4\rho - 1}, \quad y(\rho) = \frac{4\rho}{4\rho - 1}$$

  which is a local as well as a global minimizer

Penalty function methods for equally constrained problems. Example (cont.)

- Note that $\boldsymbol{x}(\rho)$ is infeasible to the original constrained problem, since

$$h(\boldsymbol{x}(\rho)) = x(\rho) + 2y(\rho) - 4 = \frac{16\rho}{4\rho - 1} - 4 = \frac{4}{4\rho - 1} \neq 0$$

- At any solution $\boldsymbol{x}(\rho)$ we can define a Lagrange multiplier estimate as

$$\lambda(\rho) = -\rho h(\boldsymbol{x}(\rho)) = -\frac{4\rho}{4\rho - 1}$$

- As $\rho \to \infty$, we get

$$\lim_{\rho\to\infty} x(\rho) = \lim_{\rho\to\infty} \frac{2}{1 - 1/(4\rho)} = 2, \quad \lim_{\rho\to\infty} y(\rho) = \lim_{\rho\to\infty} \frac{1}{1 - 1/(4\rho)} = 1$$

and indeed $\boldsymbol{x}^* = (2, 1)^T$ is the solution for the constrained problem

- Furthermore

$$\lim_{\rho\to\infty} \lambda(\rho) = \lim_{\rho\to\infty} -\frac{1}{1 - 1/(4\rho)} = -1$$

and indeed $\lambda^* = -1$ is the Lagrange multiplier at $\boldsymbol{x}^*$

Penalty function methods for equally constrained problems

Penalty functions for equally constrained problems suffer from the same problems of ill-conditioning as do barrier functions

- As the penalty parameter increases, the condition number of the Hessian matrix of $\tilde{f}_\rho(\boldsymbol{x}(\rho))$ increases, tending to ∞ as $\rho \to \infty$
- Therefore, the unconstrained minimization problems can become increasingly difficult (or even impossible) to solve

Penalty function methods for equally constrained problems. Example (cont.)

- In the above example (page 48), to demonstrate the ill-conditioning of the penalty function, we compute its Hessian matrix at $\boldsymbol{x}(\rho)$

$$\nabla^2_{\boldsymbol{xx}} f_\rho(\boldsymbol{x}) = \begin{pmatrix} \rho & 2\rho - 1 \\ 2\rho - 1 & 4\rho \end{pmatrix}$$

  It can be shown that its condition number is approximately $25\rho/4$. When ρ is large, the Hessian matrix is ill conditioned

4. Penalty function methods for inequality constrained problems

Penalty function methods for inequality constrained problems

Consider the inequality constrained problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \geq 0, \quad i = 1, \dots, p \end{array}$$

- Any continuous function $\phi(\boldsymbol{x})$ that satisfies the conditions

$$\begin{array}{ll} \phi(\boldsymbol{x}) = 0 & \text{if } \boldsymbol{x} \text{ is feasible} \\ \phi(\boldsymbol{x}) > 0 & \text{otherwise} \end{array}$$

  can serve as a penalty

- The quadratic-loss function is defined as

$$\phi(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{p}[\min(g_i(\boldsymbol{x}), 0)]^2$$

- This function has continuous first derivatives

$$\nabla\phi(\boldsymbol{x}) = \sum_{i=1}^{p}[\min(g_i(\boldsymbol{x}), 0)]\nabla g_i(\boldsymbol{x})$$

  but its second derivatives can be discontinuous at points where some constraint g_i is satisfied exactly

5. Convergence

Convergence

We focus on the convergence of barrier methods when applied to the inequality-constrained problem

Convergence results for penalty methods can be developed in a similar manner

Consider the problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \geq 0, \quad i = 1, \dots, p \end{array}$$

with the following general assumptions

1. The functions f, $g_1,\dots,g_p$ are **continuous** in $\mathbb{R}^n$
2. The set $X^\alpha = \{\boldsymbol{x} \mid \boldsymbol{x} \in X,\ f(\boldsymbol{x}) \leq \alpha\}$ is **bounded** for any finite α
3. The **interior** of X, $\mathring{X} = \{\boldsymbol{x} \mid g_i(\boldsymbol{x}) > 0, \quad i = 1, \dots, p\}$, is nonempty
4. X is the **closure** of $\mathring{X}$

Convergence

1. **Assumptions 1 and 2** imply that the function f has a minimum value on the set X. We denote this minimum value by f^*
2. **Assumption 3** is necessary to define the barrier subproblems
3. **Assumption 4** is necessary to avoid situations where the minimum point is isolated and does not have neighboring interior points

As an example of this fact, consider the problem

$$\min\ x \quad \text{subject to} \quad x^2 - 1 \geq 0, \quad x + 1 \geq 0.$$

Since $x^2 - 1 \geq 0 \Leftrightarrow |x| \geq 1$, and $x + 1 \geq 0 \Leftrightarrow x \geq -1$, the feasible set is

$$X = \{x \mid x \geq 1\} \cup \{x = -1\}$$

The point $x = -1$ is the minimizer, but because it is isolated it is not possible to approach it from the interior of the feasible region, and a barrier method could not converge to this solution

Convergence

The barrier function will be of the form

$$\tilde{f}_\mu(\boldsymbol{x}) = f(\boldsymbol{x}) + \mu\phi(\boldsymbol{x})$$

where $\phi(\boldsymbol{x})$ can be any function that is continuous on the interior of the feasible set, and that satisfies

$$\text{if} \quad \boldsymbol{x} \to \partial X \quad \text{with} \quad \boldsymbol{x} \in \mathring{X}, \quad \text{then} \quad \phi(\boldsymbol{x}) \to \infty$$

Note that if $\boldsymbol{x} \to \partial X$ with $\boldsymbol{x} \in \mathring{X}$, then $g_i(\boldsymbol{x}) \to 0_+$

- We will show here that under mild conditions, **the sequence of barrier minimizers has a convergent subsequence**, and **the limit of any such convergent subsequence is a solution to the problem**
- Although in practice, convergence of the entire sequence of minimizers is observed, from a theoretical point of view it is not always possible to guarantee convergence of the entire sequence, but only convergence of some subsequence, as will be shown in the following example

Convergence. Example

Consider the problem

$$\begin{array}{ll} \min & f(x) = -x^2 \\ \text{subject to} & g(x) = 1 - x^2 \geq 0 \end{array}$$

The logarithmic barrier function is

$$\tilde{f}_\mu(x) = -x^2 - \mu \log(1 - x^2)$$

- If $\mu \geq 1$, the problem has a single minimizer $x = 0$
- If $\mu < 1$, there are two minimizers $x = \pm\sqrt{1-\mu}$
- (The point $x = 0$ is a local maximizer if $\mu < 1$)

Suppose that $\{\mu_k\}$ is a sequence of decreasing barrier parameters less than 1

- Then, a possible sequence of minimizers of $\tilde{f}_\mu(x)$ is $x_k = (-1)^k\sqrt{1-\mu_k}$
- This sequence oscillates between neighborhoods of -1 and $+1$, and hence is nonconvergent
- However, the subsequences $\{x_{2k}\}$ and $\{x_{2k+1}\}$ both converge to solutions of the original constrained problem

Convergence theorem

Theorem

Suppose that

- *The nonlinear inequality-constrained problem satisfies conditions 1.,...,4. above*
- *A sequence of unconstrained minimization problems*

$$\min \quad \tilde{f}_\mu(\boldsymbol{x}) = f(\boldsymbol{x}) + \mu\phi(\boldsymbol{x})$$

  is solved for μ taking values $\mu_1 > \mu_2 > \cdots > \mu_k > \cdots$, where $\lim_{k\to\infty} \mu_k = 0$
- *The functions $\tilde{f}_{\mu_k}(\boldsymbol{x})$ have a minimum in $\mathring{X}$ for each k, and let x_k denote a global minimizer of $\tilde{f}_{\mu_k}(\boldsymbol{x})$*

Then

1. $f(x_{k+1}) \leq f(\boldsymbol{x}_k)$
2. $\phi(\boldsymbol{x}_{k+1}) \geq \phi(\boldsymbol{x}_k)$
3. *The sequence $\{\boldsymbol{x}_k\}$ has a convergent subsequence $\{\boldsymbol{x}_{k_j}\}$*
4. *If $\{\boldsymbol{x}_{k_j}\}$ is any convergent subsequence of unconstrained minimizers of $\tilde{f}_{\mu_k}(\boldsymbol{x})$, then its limit point is a global solution of the constrained problem*

Proof:

1. Since x_k is the minimizer of $\tilde{f}_{\mu_k}(\boldsymbol{x})$, then $\tilde{f}_{\mu_k}(\boldsymbol{x}_k) \leq \tilde{f}_{\mu_{k+1}}(\boldsymbol{x}_{k+1})$, so

$$f(\boldsymbol{x}_k) + \mu_k \phi(\boldsymbol{x}_k) \leq f(\boldsymbol{x}_{k+1}) + \mu_{k+1}\phi(\boldsymbol{x}_{k+1})$$

Also, since x_{k+1} is the minimizer of $\tilde{f}_{\mu_{k+1}}(\boldsymbol{x})$, then $\tilde{f}_{\mu_{k+1}}(\boldsymbol{x}_{k+1}) \leq \tilde{f}_{\mu_{k+1}}(\boldsymbol{x}_k)$, so

$$f(\boldsymbol{x}_{k+1}) + \mu_{k+1}\phi(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k) + \mu_{k+1}\phi(\boldsymbol{x}_k)$$

Multiplying the first inequality byμ_{k+1}, the second inequality by μ_k, adding the resulting inequalities, and reordering yields

$$(\mu_k - \mu_{k+1}) f(\boldsymbol{x}_{k+1}) \leq (\mu_k - \mu_{k+1}) f(\boldsymbol{x}_k)$$

Since $\mu_k > \mu_{k+1}$, we conclude that

$$f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k)$$

Convergence theorem. Proof (cont.)

2. As before, since x_k is the minimizer of $\tilde{f}_{\mu_k}(\mathbf{x})$, then

$$f(\mathbf{x}_k) + \mu_k \phi(\mathbf{x}_k) \leq f(\mathbf{x}_{k+1}) + \mu_{k+1} \phi(\mathbf{x}_{k+1})$$

   Since

$$f(\mathbf{x}_k) \geq f(\mathbf{x}_{k+1})$$

   this implies that

$$\phi(\mathbf{x}_{k+1}) \geq \phi(\mathbf{x}_k)$$

3. Consider the set $X^1 = \{\mathbf{x} \in X \mid f(\mathbf{x}) \leq f(\mathbf{x}_1)\}$

   The continuity of f implies that X^1 is closed, and the general assumption 2. implies that it is bounded, hence X^1 is compact

   Now, in view of 1, $f(\mathbf{x}_k) \leq f(\mathbf{x}_1)$ for all k, thus the sequence $\{x_k\}$ lies in the compact set X^1

   Therefore, $\{x_k\}$ has a convergent subsequence in X^1, and thus also in X

Convergence theorem. Proof (cont.)

4. Let $\{\boldsymbol{x}_{k_j}\}$ be a convergent subsequence of $\{\boldsymbol{x}_k\}$, and let $\hat{\boldsymbol{x}}$ be its limit point. Since $g_i(\boldsymbol{x}_{k_j}) > 0$ for all k_j, $g_i(\hat{\boldsymbol{x}}) \geq 0$, and hence $\hat{\boldsymbol{x}}$ is feasible to the constrained problem

   Let f^* be the minimum value of f in the feasible region X. We will show that $f(\hat{\boldsymbol{x}}) = f^*$ by contradiction

   Assume that

   $$f(\hat{\boldsymbol{x}}) > f^*$$

   then, from the general assumption 4. it follows that there exists some **strictly feasible** point $\boldsymbol{y} \in \mathring{X}$ such that

   $$f(\boldsymbol{y}) < f(\hat{\boldsymbol{x}}) \tag{1}$$

   Define $\epsilon = f(\hat{\boldsymbol{x}}) - f(\boldsymbol{y}) > 0$. Because f is continuous, it holds that

   $$\lim_{k_j \to \infty} f(\boldsymbol{x}_{k_j}) = f(\hat{\boldsymbol{x}})$$

   and thus for sufficiently large k_j we have

   $$f(\boldsymbol{y}) + \frac{1}{2}\epsilon < f(\boldsymbol{x}_{k_j}) \tag{2}$$

   Also, because $\boldsymbol{x}_{k_j}$ is a minimizer of $\tilde{f}_{\mu_{k_j}}(\boldsymbol{x})$ we have

   $$f(\boldsymbol{x}_{k_j}) + \mu_{k_j}\phi(\boldsymbol{x}_{k_j}) \leq f(\boldsymbol{y}) + \mu_{k_j}\phi(\boldsymbol{y}) \tag{3}$$

Convergence theorem. Proof (cont.)

We consider two cases:

- $\hat{x}$ is strictly feasible. Then, for k_j large enough , $\boldsymbol{x}_{k_j}$ is strictly feasible and, therefore, $\phi(\boldsymbol{x}_{k_j})$ is bounded. Also, because y is strictly feasible, $\phi(\boldsymbol{y})$ is bounded.

  Therefore, for k_j sufficiently large

  $$-\frac{1}{8}\epsilon \le \mu_{k_j}\phi(\boldsymbol{x}_{k_j}) \quad \text{and} \quad \mu_{k_j}\phi(\boldsymbol{y}) \le \frac{1}{8}\epsilon$$

  Combining this with (3) yields

  $$f(\boldsymbol{x}_{k_j}) - \frac{1}{8}\epsilon \le f(\boldsymbol{y}) + \frac{1}{8}\epsilon \quad \Rightarrow \quad f(\boldsymbol{x}_{k_j}) \le f(\boldsymbol{y}) + \frac{1}{4}\epsilon$$

  But this is a contradiction to (2) and, therefore, to (1)

- $\hat{x}$ is not strictly feasible. It follows from (2) that $f(\boldsymbol{y}) < f(\boldsymbol{x}_{k_j})$.

  Adding this to (3), rearranging, and dividing by μ_{k_j} gives $\phi(\boldsymbol{x}_{k_j}) < \phi(\boldsymbol{y})$.

  Because $\boldsymbol{y}$ is strictly feasible, the right-hand side is finite. Nevertheless, because $\boldsymbol{x}_{k_j}$ approaches the boundary, the left-hand side is unbounded above as k_j tends to ∞. Therefore have a contradiction to (1)

6. Stabilized penalty and barrier methods

Stabilized penalty and barrier methods

Despite the ill-conditioning of the Hessian matrix of the barrier function, it is possible to compute a Newton-type direction in a numerically stable manner

Consider again the problem

$$\begin{array}{ll} \text{min} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \geq 0 \end{array}$$

We will use the logarithmic barrier function, but **the results can be extended to other penalty and barrier methods**

Stabilized penalty and barrier methods

- Let $A = \nabla \boldsymbol{g}(\boldsymbol{x})^T$ be the Jacobian matrix of the constraints, and assume that A has full rank
- Let Z be a basis matrix for the null space of A, and let A_r be a right-inverse matrix for A, this is: $AA_r = I$
- We assume that Z and A_r have been obtained from an orthogonal QR factorization of A, so that

$$\begin{pmatrix} Z & A^T \end{pmatrix} \begin{pmatrix} Z^T \\ A_r^T \end{pmatrix} = Id$$

- We also define the Lagrange multiplier estimates $\lambda_i = \mu / g_i(\boldsymbol{x})$ and the diagonal matrix D, whose i-th diagonal entry is λ_i

Stabilized penalty and barrier methods

- Let $B = \nabla^2_{\boldsymbol{xx}} \tilde{f}_\mu(\boldsymbol{x})$ be the Hessian of the barrier function

$$B = \nabla^2_{\boldsymbol{xx}} \tilde{f}_\mu(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}) - \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(\boldsymbol{x}) + \frac{1}{\mu} \sum_{i=1}^{p} \lambda_i^2 \nabla g_i(\boldsymbol{x})(\nabla g_i(\boldsymbol{x}))^T =$$

$$= H + \frac{1}{\mu} A^T D A$$

where

$$H = \nabla^2 f(\boldsymbol{x}) - \sum_{i=1}^{p} \lambda_i \nabla^2 g_i(\boldsymbol{x})$$

is the Hessian matrix of the Lagrangian

- From the identity

$$B = I\,B\,I = \begin{pmatrix} Z & A^T \end{pmatrix} \begin{pmatrix} Z^T \\ A_r^T \end{pmatrix} B \begin{pmatrix} Z & A_r \end{pmatrix} \begin{pmatrix} Z^T \\ A \end{pmatrix}$$

we get

$$B^{-1} = \begin{pmatrix} Z & A_r \end{pmatrix} \begin{pmatrix} Z^T B Z & Z^T B A_r \\ A_r^T B Z & A_r^T B A_r \end{pmatrix}^{-1} \begin{pmatrix} Z^T \\ A_r^T \end{pmatrix}$$

Stabilized penalty and barrier methods

To compute the search direction $\boldsymbol{z} = B^{-1}\nabla_{\boldsymbol{x}}\tilde{f}_\mu(\boldsymbol{x})$ of Newton's method, we approximate B^{-1} using the **bordered-inverse formula**

If A_1 and A_3 are symmetric matrices, then

$$\begin{pmatrix} A_1 & A_2 \\ A_2^T & A_3 \end{pmatrix}^{-1} = \begin{pmatrix} A_1^{-1} + A_1^{-1}A_2G^{-1}A_2^TA_1^{-1} & -G^{-1}A_2^TA_1^{-1} \\ -A_1^{-1}A_2G^{-1} & G^{-1} \end{pmatrix} =$$

$$= \begin{pmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} A_1^{-1}A_2 \\ -I \end{pmatrix} G^{-1} \begin{pmatrix} A_2^TA_1^{-1} & -I \end{pmatrix}$$

Applying this formula, and noting that $AZ = 0$, gives

$$B^{-1} = \begin{pmatrix} Z & A_r \end{pmatrix} \left[\begin{pmatrix} (Z^THZ)^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \right.$$

$$\left. + \begin{pmatrix} (Z^THZ)^{-1}Z^THA_r \\ -I \end{pmatrix} G^{-1} \begin{pmatrix} A_r^THZ(Z^THZ)^{-1} & -I \end{pmatrix} \right] \begin{pmatrix} (Z^T \\ A_r^T \end{pmatrix}$$

where

$$G = \frac{1}{\mu}D + A_r^THA_r - A_r^THZ(Z^THZ)^{-1}Z^THA_r$$

Stabilized penalty and barrier methods

When μ is small (that is, as we approach the solution where the ill-conditioning becomes apparent), $G^{-1} \approx \mu D^{-1}$. Hence

$$B^{-1} \approx Z(Z^T HZ)^{-1} Z^T + \mu(Z(Z^T HZ)^{-1} Z^T H - I) A_r D^{-1} A_r^T (HZ(Z^T HZ)^{-1} Z^T - I)$$

This approximation to B^{-1} determines an approximation to the Newton direction

$$\mathbf{z} = B^{-1} \nabla_{\mathbf{x}} \tilde{f}_\mu(\mathbf{x}) \approx \mathbf{z}_1 + \mu \mathbf{z}_2$$

where

$$\begin{aligned} \mathbf{z}_1 &= -Z(Z^T HZ)^{-1} Z^T \nabla_{\mathbf{x}} \tilde{f}_\mu(\mathbf{x}) \\ \lambda &= A_r^T (H\mathbf{z}_1 + \nabla_{\mathbf{x}} \tilde{f}_\mu(\mathbf{x})) \\ \mathbf{z}_2 &= (Z(Z^T HZ)^{-1} Z^T H - I) A_r D^{-1} \lambda \end{aligned}$$

As $\mu \rightarrow 0$, it can be shown that the error in the approximate search direction is $O(\mu)$

Stabilized penalty and barrier methods. Example

Example: Solve

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = x^2 + y^2 \\ \text{subject to} & g(x, y) = x - 1 \geq 0 \end{array}$$

The associated barrier problem is

$$\min \quad \tilde{f}_\mu(\boldsymbol{x}) = x^2 + y^2 - \mu \log(x - 1)$$

If we set $\mu = 10^{-4}$ and $\boldsymbol{x} = (1.001,\ 0.001)^T$, then the multiplier estimate is $\lambda = \mu/(x - 1) = 0.1$. At this point

$$\nabla_{\boldsymbol{x}} \tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} \frac{2x-\mu}{x-1} \\ 2y \end{pmatrix} = \begin{pmatrix} 1.902 \\ 0.002 \end{pmatrix}$$

$$B = \nabla^2_{\boldsymbol{xx}} \tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} 2 + \frac{2x-\mu}{(x-1)^2} & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 102 & 0 \\ 0 & 2 \end{pmatrix}$$

It follows that $\operatorname{cond}(B) = 50.1$, and the Newton direction is

$$\boldsymbol{z} = B^{-1} \nabla_{\boldsymbol{x}} \tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} -0.0186 \\ -0.0010 \end{pmatrix}$$

Stabilized penalty and barrier methods. Example (cont.)

We now determine the approximate Newton direction

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad A_r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad Z^T HZ = (2)$$

$$\operatorname{cond}(B) = 1, \quad D = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$$

From these we determine

$$\begin{aligned} \boldsymbol{z}_1 &= -Z(Z^T HZ)^{-1} Z^T \nabla_{\boldsymbol{x}} \tilde{f}_\mu(\boldsymbol{x}) = \begin{pmatrix} 0 \\ -0.001 \end{pmatrix} \\ \lambda &= A_r^T (H\boldsymbol{z}_1 + \nabla_{\boldsymbol{x}} \tilde{f}_\mu(\boldsymbol{x})) = 1.902 \\ \boldsymbol{z}_2 &= (Z(Z^T HZ)^{-1} Z^T H - I) A_r D^{-1} \lambda = \begin{pmatrix} -190.2 \\ 0 \end{pmatrix} \end{aligned}$$

and the approximate Newton direction is

$$\bar{\boldsymbol{z}} = \boldsymbol{z}_1 + \mu \boldsymbol{z}_2 = \begin{pmatrix} -0.0190 \\ -0.0010 \end{pmatrix}$$

which is very close to the Newton direction previously computed (error $\approx 10^{-4}$)