# TMA4180 Optimization: Augmented Lagrangian Method; Logarithmic Barrier Methods

Elisabeth Köbis

NTNU, TMA4180, Optimering, Spring 2021

March 15th, 2021

Remember to register your visit in S5!

# Plan for the Day

1. Equality-Constrained Problems: Augmented Lagrangian method (N&W, Chapter 17.3)
2. Inequality-Constrained Problems: Logarithmic Barrier Methods (N&W, Chapter 19.6)

# Equality-Constrained Problems: Augmented Lagrangian method (N&W, 17.3)

# Problem and Quadratic Penalty Function
## Equality-Constrained Problems

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to } c_i(x) = 0, \quad i \in \mathcal{E}. \tag{P}$$

Let $\mu > 0$ be a *penalty parameter*. We define

$$\min_{x\in\mathbb{R}^n} \; Q(x;\mu) := f(x) + \frac{\mu}{2}\sum_{i\in\mathcal{E}} c_i^2(x). \tag{P$_{\text{penalty}}$}$$

By letting $\mu \to +\infty$, we penalize the constraint violations with increasing severity. Recall that the composition $g \circ c$ (here: $g \circ c = c^2$) of smooth functions defined on open domains of Euclidean spaces is smooth, and so are its sums. Therefore, if $f, c_i$ $(i \in \mathcal{E})$ are smooth, $Q(\cdot, \mu)$ is smooth, and we can use techniques from unconstrained optimization to search for a minimizer of (P$_{\text{penalty}}$).

**The approximate minimizers $x_k$ of $Q(x;\mu_k)$ do not quite satisfy the feasibility conditions $c_i(x) = 0$, $i \in \mathcal{E}$.**

# Quadratic Penalty Function Equality-Constrained Problems

Instead, they are perturbed, so that

$$\boxed{c_i(x_k) \approx -\lambda_i^*/\mu_k \text{ for all } i \in \mathcal{E}} \tag{1}$$

(see Theorem 17.1 in N&W, or see the next slide).

# Convergence of the Quadratic Penalty Method for Equality-Constrained Problems

## Theorem

Suppose that $\mathcal{I} = \emptyset$ and that the tolerances and penalty parameters satisfy $\tau_k \to 0$ and $\mu_k \to +\infty$. Then if a limit point $x^*$ of the sequence $\{x_k\}$ is infeasible, it is a stationary point of the function $||c(x)||^2$. On the other hand, if a limit point $x^*$ is feasible and the constraint gradients $\nabla c_i(x^*)$ are linearly independent, then $x^*$ is a KKT point for the problem (P). For such points, we have for any infinite subsequence $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} x_k = x^*$ that

$$\boxed{\forall i \in \mathcal{E} : \quad \lim_{k\in\mathcal{K}} -\mu_k c_i(x_k) = \lambda_i^*,}$$

where $\lambda^*$ is the multiplier vector that satisfies the KKT conditions for the equality constrained problem.

# Quadratic Penalty Function Equality-Constrained Problems

We have $c_i(x_k) \to 0$ as $\mu_k \to +\infty$, but one may ask whether we can alter the function $Q(x; \mu_k)$ to avoid this systematic perturbation–that is, to make the approximate minimizers more nearly satisfy the equality constraints $c_i(x) = 0$, even for moderate values of $\mu_k$.

$\longrightarrow$ Augmented Lagrangian achieves this goal.

# Equality-Constrained Problem, Augmented Lagrangian Method

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to } c_i(x) = 0, \quad i \in \mathcal{E}. \tag{P}$$

## Augmented Lagrangian

$$\mathcal{L}_A(x, \lambda; \mu) := f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x),$$

where $\lambda_i$ $(i \in \mathcal{E})$ are estimate of the Lagrange multipliers of (P).

## Remark

The augmented Lagrangian differs from the (standard) Lagrangian for by the presence of the squared terms, while it differs from the quadratic penalty function in the presence of the summation term involving $\lambda$. In this sense, it is a combination of the Lagrangian function and the quadratic penalty function.

# Equality-Constrained Problem, Augmented Lagrangian Method

Fix the penalty parameter $\mu_k > 0$ at the $k$th iteration, fix $\lambda$ at the current estimate $\lambda^k$, and perform minimization of $\mathcal{L}_A$ w.r.t. $x$. By the optimality condition for unconstrained optimization, we have

$$0 \approx \nabla_x \mathcal{L}_A(x_k, \lambda^k; \mu_k) = \nabla f(x_k) - \sum_{i \in \mathcal{E}} \underbrace{\left[\lambda_i^k - \mu_k c_i(x_k)\right]}_{=\lambda_i^*} \nabla c_i(x_k).$$

By the optimality condition (KKT) for problem (P)[1], we get

$$\text{for all } i \in \mathcal{E}: \ \lambda_i^* \approx \lambda_i^k - \mu_k c_i(x_k). \tag{2}$$

Rearranging yields

$$\text{for all } i \in \mathcal{E}: \ c_i(x_k) \approx -\frac{1}{\mu_k}(\lambda_i^* - \lambda_i^k).$$

[1] Theorem (First-Order Necessary Optimality Conditions): Suppose that $x^*$ is a local minimizer of (P), that the functions $f$ and $c_i$ in are continuously differentiable, and that the LICQ holds at $x^*$. Then there is a Lagrange multiplier vector $\lambda^*$ with components $\lambda_i^*$, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at $(x^*, \lambda^*)$:

$$\boxed{\nabla_x \mathcal{L}(x^*, \lambda^*) = 0}, \ c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E}, \ c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I},$$

$$\lambda_i \geq 0, \quad \text{for all } i \in \mathcal{I}, \ \lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.$$

# Equality-Constrained Problem, Augmented Lagrangian Method

Conclusion: If $\lambda^k$ is close to $\lambda^*$, the infeasibility of $x_k$ will be much smaller than $\frac{1}{\mu_k}$, rather than being proportional to $\frac{1}{\mu_k}$ as in (1).

The relation (2) immediately suggests a formula for improving our current estimate $\lambda^k$ of the Lagrange multiplier vector, using the approximate minimizer $x_k$ just calculated: We can set

$$\text{for all } i \in \mathcal{E} : \; \lambda_i^{k+1} \approx \lambda_i^k - \mu_k c_i(x_k). \tag{3}$$

# Augmented Lagrangian Method

## Algorithm: Augmented Lagrangian Method - Equality Constraints

Input: $\mu_0 > 0$, tolerance $\tau_0 > 0$, starting points $x_0^s$ and $\lambda^0$

**for** $k = 0, 1, 2, \ldots$

Find an approximate minimizer $x_k$ of $\mathcal{L}(\cdot, \lambda^k; \mu_k)$, starting at $x_k^s$, and terminating when $||\nabla_x \mathcal{L}_A(x_k, \lambda^k; \mu_k)|| \leq \tau_k$;

**if** a convergence test for (P) is satisfied

**stop** with approximate solution $x_k$;

**end if**

Update Lagrange multipliers using (3) to obtain $\lambda^{k+1}$

Choose new penalty parameter $\mu_{k+1} > \mu_k$;

Choose new starting point for the next iteration $x_{k+1}^s = x_k$;

Select tolerance $\tau_{k+1}$

**end for**

# Augmented Lagrangian vs. Quadratic Penalty Function

## Example

$$\min_{x \in \mathbb{R}^n} x_1 + x_2 \quad \text{subject to } x_1^2 + x_2^2 - 2 = 0$$

with the minimizer $(-1, -1)^T$ and optimal Lagrange multiplier $\lambda^* = -0.5$. The quadratic penalty function is

$$Q(x; \mu) = x_1 + x_2 + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2,$$

while the augmented Lagrangian is

$$\mathcal{L}_A(x, \lambda; \mu) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2) + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2.$$

| | $\mu = 2$ | $\mu = 20$ | $\mu = 200$ |
|---|---|---|---|
| Quadr. Pen. | $x = (-1.057, -1.057)^T$ | $x = (-1.006, -1.006)^T$ | $x = (-1.0006, -1.0006)^T$ |
| Aug. Lagr. ($\lambda = -0.4$) | $x = (-1.012, -1.012)^T$ | $x = (-1.001, -1.001)^T$ | $x = (-1.00012, -1.00012)^T$ |
| Aug. Lagr. ($\lambda = -0.5$) | $x = (-1, -1)^T$ | $x = (-1, -1)^T$ | $x = (-1, -1)^T$ |

# Augmented Lagrangian: Properties

### Theorem

Let $x^*$ be a local solution of (P) at which the LICQ is satisfied (that is, the gradients $\nabla c_i(x^*)$, $i \in \mathcal{E}$, are linearly independent vectors), and the second-order sufficient conditions specified in Theorem 12.6 in N&W are satisfied for $\lambda = \lambda^*$. Then there is a threshold value $\overline{\mu}$ such that for all $\mu \geq \overline{\mu}$, $x^*$ is a strict local minimizer of $\mathcal{L}_A(x, \lambda^*; \mu)$.

Proof:

# Augmented Lagrangian: Properties

### Remark

Although we do not know $\lambda^*$ exactly in practice, the result and its proof suggest that we can obtain a good estimate of $x^*$ by minimizing $\mathcal{L}_A(x, \lambda^*; \mu)$ even when $\mu$ is not particularly large, provided that $\lambda$ is a reasonably good estimate of $\lambda^*$.

# Augmented Lagrangian: Properties

## Theorem

Suppose that the assumptions of the preceding theorem are satisfied at $x^*$ and $\lambda^*$ and let $\overline{\mu}$ be chosen as in that theorem. Then there exist positive scalars $\delta$, $\epsilon$ and $M$ such that the following claims hold:

(a) For all $\lambda^k$ and $\mu_k$ satisfying

$$||\lambda^k - \lambda^*|| \leq \mu_k \delta, \quad \mu_k \geq \overline{\mu}, \tag{4}$$

the problem

$$\min_x \mathcal{L}_A(x, \lambda^k; \mu_k) \quad \text{subject to } ||x - x^*|| \leq \epsilon$$

has a unique solution $x_k$. Moreover, we have

$$\underbrace{||x_k - x^*|| \leq M||\lambda^k - \lambda^*||/\mu_k.}_{x_k \text{ will be close to } x^* \text{ if } \lambda^k \text{ is accurate } \textit{or} \text{ if the penalty parameter } \mu_k \text{ is large}}$$

# Augmented Lagrangian: Properties

## Theorem, continued

(b) For all $\lambda^k$ and $\mu_k$ that satisfy (4), we have

$$\underbrace{||\lambda^{k+1} - \lambda^*|| \leq M||\lambda^k - \lambda^*||/\mu_k,}_{\text{locally, we can ensure an improvement in the accuracy of the multipliers by choosing a sufficiently large value of } \mu_k}$$

where $\lambda^{k+1}$ is given by $\lambda_i^{k+1} = \lambda_i^k - \mu_k c_i(x_k)$ for all $i \in \mathcal{E}$.

(c) For all $\lambda^k$ and $\mu_k$ that satisfy (4), the matrix $\nabla^2_{xx}\mathcal{L}_A(x_k, \lambda^k; \mu_k)$ is positive definite and the constraint gradients $\nabla c_i(x_k)$, $i \in \mathcal{E}$, are linearly independent.

$\Longrightarrow$ second-order sufficient conditions for unconstrained minimization are satisfied for the $k$th subproblem under the given conditions

Proof: See Proposition 4.2.3 in D. P. Bertsekas: Nonlinear Programming, Athena Scientific, Belmont, MA, second edition, 1999.

# Logarithmic Barrier Methods (N&W 19.6)

# Inequality-Constrained Problem: Characteristic Function

Consider the inequality-constrained problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to } c_i(x) \geq 0, \quad i \in \mathcal{I}. \tag{P}$$

(P) is equivalent[2] to

$$\min_{x \in \mathbb{R}^n} f(x) + \chi_\Omega(x),$$

where $\Omega = \{x \in \mathbb{R}^n | c_i(x) \geq 0,\ i \in \mathcal{I}\}$ and $\chi_\Omega$ is the *characteristic function* (also called *indicator function*) from convex analysis:

$$\chi_\Omega(x) := \begin{cases} 0 & (x \in \Omega) \\ +\infty & (\text{else}) \end{cases}$$

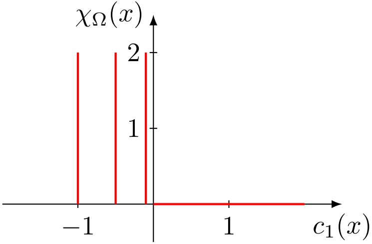
Problem: $f(x) + \chi_\Omega(x)$ is not continuous.

[2] in the sense that the minimizers of both problems coincide

# Inequality-Constrained Problem Characteristic Function

Let $\mathcal{I} = \{1\}$. Then $\chi_\Omega$ is:

$$\chi_\Omega(x) = \begin{cases} 0 & (c_1(x) \geq 0) \\ +\infty & (\text{else}) \end{cases}$$

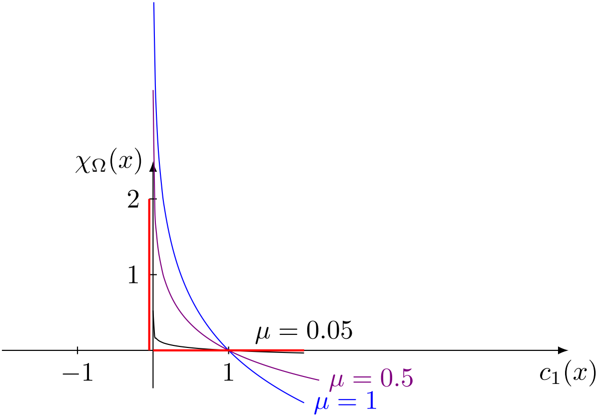


# Logarithmic Barrier Function

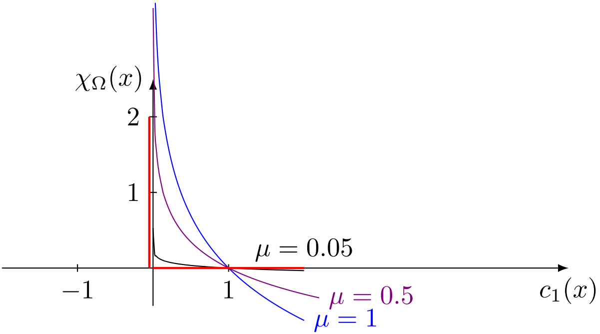
For $\mu > 0$, we define the *log-barrier function* as

$$P(x;\mu) = f(x) - \mu \sum_{i \in \mathcal{I}} \log c_i(x).$$

For $\mu \to 0$ (from above), $-\mu \log c_1(x)$ is a smooth approximation of $\chi_\Omega$:



# Logarithmic Barrier Function




As $\mu$ approaches $0$, the approximation becomes closer to the characteristic function, as shown in the figure above. Also, for any value of $\mu$, if any of the constraints is violated, the value of the barrier approaches infinity. For feasible $x$ away from the boundary, the barrier term is small, while it is large for feasible $x$ close to the boundary (i.e., if $c(x) \approx 0$).

# Logarithmic Barrier Function

But: (P) and

$$\min_{x\in\mathbb{R}^n} P(x;\mu) = f(x) - \mu \sum_{i\in\mathcal{I}} \log c_i(x)$$

are **not** equivalent.

## Remark

It can be seen that the domain of the log barrier is the set of strictly feasible points

$$\{x \in \mathbb{R}^n | c_i(x) > 0,\ i \in \mathcal{I}\},$$

because $\log(0)$ is not defined.

# Inequality-Constrained Problem: Example

## Example

$$\min_{x \in \mathbb{R}^n} x_1^2 + x_2^2 \quad \text{s.t. } x_1 + x_2 \geq 4.$$

The logarithmic barrier function is $P(x;\mu) = x_1^2 + x_2^2 - \mu \ln(x_1 + x_2 - 4)$.
Finding a local minimizer of $P(x;\mu)$:

$$\nabla P(x;\mu) = \begin{pmatrix} 2x_1 - \mu \frac{1}{x_1+x_2-4} \\ 2x_2 - \mu \frac{1}{x_1+x_2-4} \end{pmatrix} \stackrel{!}{=} 0.$$

The stationary points are

$$x(\mu) = \begin{pmatrix} 1 + \frac{\sqrt{4-\mu}}{2} \\ 1 + \frac{\sqrt{4-\mu}}{2} \end{pmatrix},$$

and $\lim_{\mu \to 0} x(\mu) = (2,2)^T$, which is the global minimizer of the inequality-constrained problem.

# Logarithmic Barrier Function

One way to obtain an estimate of the Lagrange multipliers is based on differentiating $P$ to obtain

$$\nabla P(x;\mu) = \nabla f(x) - \sum_{i\in\mathcal{I}} \frac{\mu}{c_i(x)} \nabla c_i(x).$$

When $x$ is close to the minimizer $x(\mu)$ and $\mu$ is small, we see from Theorem 12.1 that the optimal Lagrange multipliers $z_i^*$, $i \in \mathcal{I}$, can be estimated as follows:

$$z_i^* = \frac{\mu}{c_i(x)}, \quad i \in \mathcal{I}. \tag{5}$$

# Unconstrained Primal Barrier Method by Means of Logarithmic Barrier Function

## Algorithm: Unconstrained Primal Barrier Method

Input: $\mu_0 > 0$, sequence $\{\tau_k\}$ with $\tau \to 0$, starting point $x_0^s$
**for** $k = 0, 1, 2, \ldots$
  Find an approximate minimizer $x_k$ of $P(\cdot; \mu_k)$, starting at $x_k^s$ and terminating when $||\nabla P(x_k; \mu_k)|| \leq \tau_k$; Compute Lagrange multipliers $z_k$ by (5)
  **if** final convergence test satisfied
    **stop** with approximate solution $x_k$;
  **end if**
  Choose new penalty parameter $\mu_{k+1} < \mu_k$;
  Choose new starting point $x_{k+1}^s$;
**end for**