Lecture 13: Augmented Lagrangian methods for constrained optimization problems

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Nonlinear equality-constrained problems (again)

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad c(x) = 0, \tag{eCP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c = (c_1, \ldots, c_m) : \mathbb{R}^n \to \mathbb{R}^m$ smooth.

Nonlinear equality-constrained problems (again)

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad c(x) = 0, \qquad \text{(eCP)}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c = (c_1, \ldots, c_m) : \mathbb{R}^n \to \mathbb{R}^m$ smooth.

Another example of merit function and method for (eCP): augmented Lagrangian function

$$\Phi(x, u, \sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma} \|c(x)\|^2$$

where both $u \in \mathbb{R}^m$ and $\sigma > 0$ are now algorithm parameters.

Nonlinear equality-constrained problems (again)

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad c(x) = 0, \qquad \text{(eCP)}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c = (c_1, \ldots, c_m) : \mathbb{R}^n \to \mathbb{R}^m$ smooth.

Another example of merit function and method for (eCP):
augmented Lagrangian function

$$\Phi(x, u, \sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma} \|c(x)\|^2$$

where both $u \in \mathbb{R}^m$ and $\sigma > 0$ are now algorithm parameters.

Two interpretations:

- shifted quadratic penalty function
- convexification of the Lagrangian function

Aim: adjust both u and σ to encourage convergence.

Derivatives of the augmented Lagrangian function

$\Phi(x, u, \sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma}\|c(x)\|^2$. Let $J(x)$ Jacobian of constraints $c(x) = (c_1(x), \ldots, c_m(x))$.

- Lagrangian: $\mathcal{L}(x, y) = f(x) - y^T c(x)$ for a(ny) $y \in \mathbb{R}^m$

Derivatives of the augmented Lagrangian function

$\Phi(x, u, \sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma}\|c(x)\|^2$. Let $J(x)$ Jacobian of constraints $c(x) = (c_1(x), \ldots, c_m(x))$.

- Lagrangian: $\mathcal{L}(x, y) = f(x) - y^T c(x)$ for a(ny) $y \in \mathbb{R}^m$
- $\nabla_x \Phi(x, u, \sigma) = \nabla f(x) - J(x)^T u + \frac{1}{\sigma} J(x)^T c(x)$

  $\Longrightarrow \quad \nabla_x \Phi(x, u, \sigma) = \nabla f(x) - J(x)^T y = \nabla_x \mathcal{L}(x, y)$

  where $y := u - \dfrac{c(x)}{\sigma}$ Lagrange multiplier estimates

Derivatives of the augmented Lagrangian function

$\Phi(x,u,\sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma}\|c(x)\|^2$. Let $J(x)$ Jacobian of constraints $c(x) = (c_1(x), \ldots, c_m(x))$.

- Lagrangian: $\mathcal{L}(x,y) = f(x) - y^T c(x)$ for a(ny) $y \in \mathbb{R}^m$
- $\nabla_x \Phi(x,u,\sigma) = \nabla f(x) - J(x)^T u + \frac{1}{\sigma} J(x)^T c(x)$
  $\Longrightarrow \quad \nabla_x \Phi(x,u,\sigma) = \nabla f(x) - J(x)^T y = \nabla_x \mathcal{L}(x,y)$
  where $y := u - \dfrac{c(x)}{\sigma}$ Lagrange multiplier estimates
- $\nabla^2 \Phi(x,u,\sigma) = \nabla^2 f(x) - \sum_{i=1}^m u_i \nabla^2 c_i(x) +$
  $\frac{1}{\sigma}\sum_{i=1}^m c_i(x)\nabla^2 c_i(x) + \frac{1}{\sigma} J(x)^T J(x)$

  $\Longrightarrow$
  $\nabla^2 \Phi(x,u,\sigma) = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 c_i(x) + \frac{1}{\sigma} J(x)^T J(x)$
  $\Longrightarrow \nabla^2 \Phi(x,u,\sigma) = \nabla^2 \mathcal{L}(x,y) + \frac{1}{\sigma} J(x)^T J(x)$ where $y = u - \frac{c(x)}{\sigma}$.

A convergence result for the augmented Lagrangian

Theorem 22. (Global convergence of augmented Lagrangian) Assume that $f, c \in \mathcal{C}^1$ in (eCP) and for $k \geq 0$, let

$$y^k = u^k - \frac{c(x^k)}{\sigma^k},$$

for some $u^k \in \mathbb{R}^m$, and assume that

$$\|\nabla\Phi(x^k, u^k, \sigma^k)\| \leq \epsilon^k, \text{ where } \epsilon^k \to 0, k \to \infty.$$

Moreover, assume that $x^k \to x^*$, where $\nabla c_i(x^*)$, $i = \overline{1, m}$, are linearly independent. Then $y^k \longrightarrow y^*$ as $k \longrightarrow \infty$ with y^* satisfying $\nabla f(x^*) - J(x^*)^T y^* = 0$.

If additionally, either $\sigma^k \to 0$ for bounded u^k or $u^k \to y^*$ for bounded σ^k then x^* is a KKT point of (eCP) with associated Lagrange multipliers y^*. $\square$

A convergence result for the augmented Lagrangian

Proof of Theorem 22. The first part of Th 22, namely, convergence of y^k to $y^* = J(x^*)^+ \nabla f(x^*)$ follows exactly as in the proof of Theorem 21 (penalty method convergence). (Note that the assumption $\sigma^k \to 0$ is not needed for this part of the proof of Th 21.)

A convergence result for the augmented Lagrangian

Proof of Theorem 22. The first part of Th 22, namely, convergence of y^k to $y^* = J(x^*)^+ \nabla f(x^*)$ follows exactly as in the proof of Theorem 21 (penalty method convergence). (Note that the assumption $\sigma^k \to 0$ is not needed for this part of the proof of Th 21.)

It remains to show that under the additional assumptions on u^k and σ^k, x^* is feasible for the constraints.

A convergence result for the augmented Lagrangian

Proof of Theorem 22. The first part of Th 22, namely, convergence of y^k to $y^* = J(x^*)^+\nabla f(x^*)$ follows exactly as in the proof of Theorem 21 (penalty method convergence). (Note that the assumption $\sigma^k \to 0$ is not needed for this part of the proof of Th 21.)

It remains to show that under the additional assumptions on u^k and σ^k, x^* is feasible for the constraints. To see this, use the definition of $y^k = u^k - c(x^k)/\sigma^k$ to deduce $c(x^k) = \sigma^k(u^k - y^k)$ and so

$$\|c(x^k)\| = \sigma^k\|u^k - y^k\| \leq \sigma^k\|y^k - y^*\| + \sigma^k\|u^k - y^*\| \quad (*)$$

On the LHS of (*): $\|c(x^k)\| \to \|c(x^*)\|$ as $x^k \to x^*$.

A convergence result for the augmented Lagrangian

Proof of Theorem 22. (continued)

$$\|c(x^k)\| = \sigma^k \|u^k - y^k\| \le \sigma^k \|y^k - y^*\| + \sigma^k \|u^k - y^*\| \quad (*)$$

A convergence result for the augmented Lagrangian

Proof of Theorem 22.(continued)

$$\|c(x^k)\| = \sigma^k \|u^k - y^k\| \leq \sigma^k \|y^k - y^*\| + \sigma^k \|u^k - y^*\| \quad (*)$$

For the RHS of (*):

(i) if $\sigma^k \to 0$ and $\|u^k\| \leq M$ for all $k \geq 0$, then by triangle ineq.,

$\sigma^k \|u^k - y^*\| \leq \sigma^k \|u^k\| + \sigma^k \|y^*\| \leq \sigma^k (M + \|y^*\|) \to 0$

Since $y^k \to y^*$, $\sigma^k \|y^k - y^*\| \to 0$ as $\{\sigma^k\}$ is bounded above.

So the RHS of (*) converges to zero as $k \to \infty$.

A convergence result for the augmented Lagrangian

Proof of Theorem 22.(continued)

$$\|c(x^k)\| = \sigma^k \|u^k - y^k\| \le \sigma^k \|y^k - y^*\| + \sigma^k \|u^k - y^*\| \quad (*)$$

For the RHS of (*):

(i) if $\sigma^k \to 0$ and $\|u^k\| \le M$ for all $k \ge 0$, then by triangle ineq.,

$\sigma^k \|u^k - y^*\| \le \sigma^k \|u^k\| + \sigma^k \|y^*\| \le \sigma^k (M + \|y^*\|) \to 0$

Since $y^k \to y^*$, $\sigma^k \|y^k - y^*\| \to 0$ as $\{\sigma^k\}$ is bounded above.

So the RHS of (*) converges to zero as $k \to \infty$.

(ii) else, if $u^k \to y^*$ and $|\sigma^k| \le \overline{\sigma}$ for all $k \ge 0$, then

$\sigma^k \|u^k - y^*\| \le \overline{\sigma} \|u^k - y^*\| \to 0.$

Since $y^k \to y^*$, $\sigma^k \|y^k - y^*\| \le \overline{\sigma} \|y^k - y^*\| \to 0.$

A convergence result for the augmented Lagrangian

Proof of Theorem 22.(continued)

$$\|c(x^k)\| = \sigma^k \|u^k - y^k\| \leq \sigma^k \|y^k - y^*\| + \sigma^k \|u^k - y^*\| \quad (*)$$

For the RHS of (*):

(i) if $\sigma^k \to 0$ and $\|u^k\| \leq M$ for all $k \geq 0$, then by triangle ineq.,

$\sigma^k \|u^k - y^*\| \leq \sigma^k \|u^k\| + \sigma^k \|y^*\| \leq \sigma^k (M + \|y^*\|) \to 0$

Since $y^k \to y^*$, $\sigma^k \|y^k - y^*\| \to 0$ as $\{\sigma^k\}$ is bounded above.

So the RHS of (*) converges to zero as $k \to \infty$.

(ii) else, if $u^k \to y^*$ and $|\sigma^k| \leq \overline{\sigma}$ for all $k \geq 0$, then

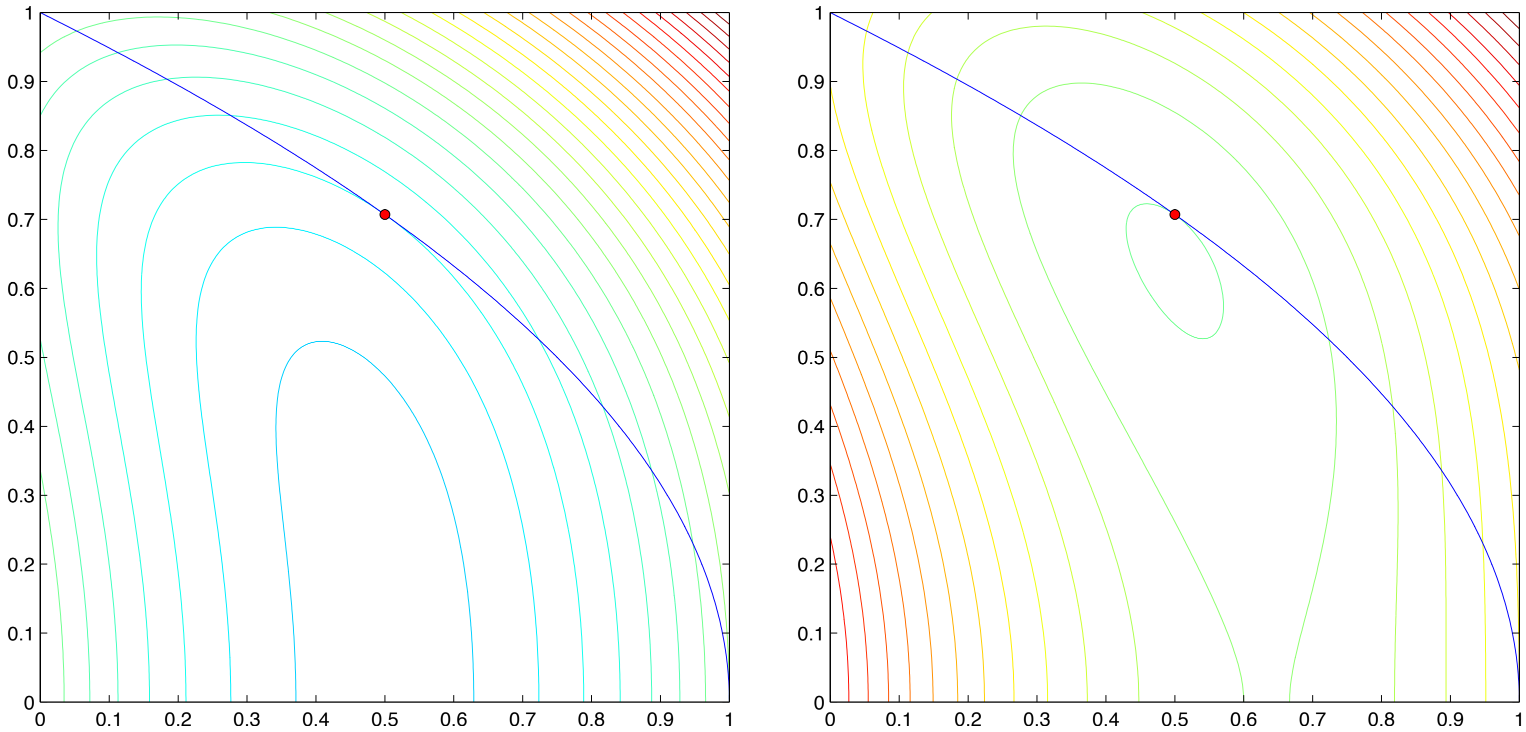
$\sigma^k \|u^k - y^*\| \leq \overline{\sigma} \|u^k - y^*\| \to 0.$

Since $y^k \to y^*$, $\sigma^k \|y^k - y^*\| \leq \overline{\sigma} \|y^k - y^*\| \to 0.$

Thus in both cases, LHS of (*) and RHS of (*) in the limit are equal and so $c(x^*) = 0$. $\square$

Note that Augmented Lagrangian may converge to KKT points without $\sigma^k \to 0$, which limits the ill-conditioning.

Contours of the augmented Lagrangian - an example

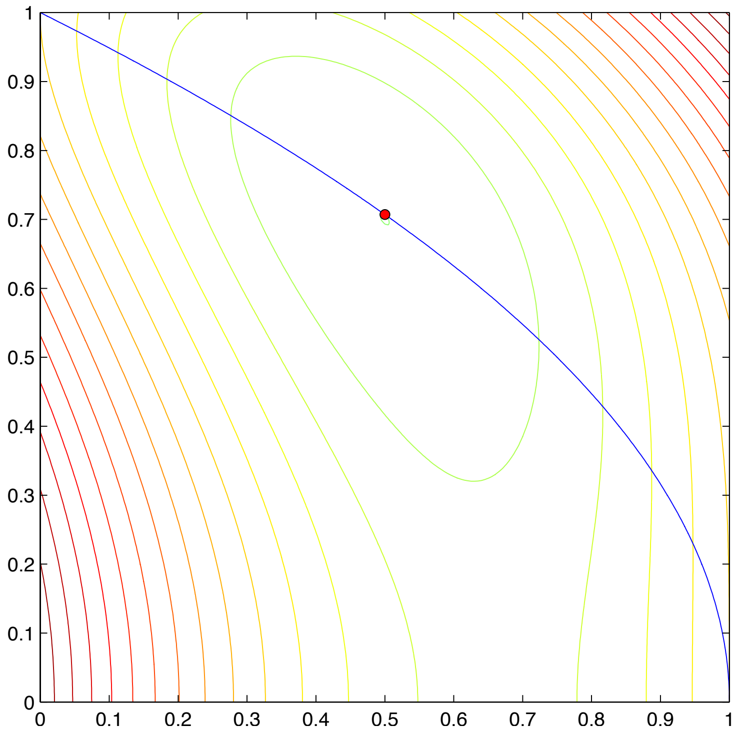


$u = 0.5$ $u = 0.9$

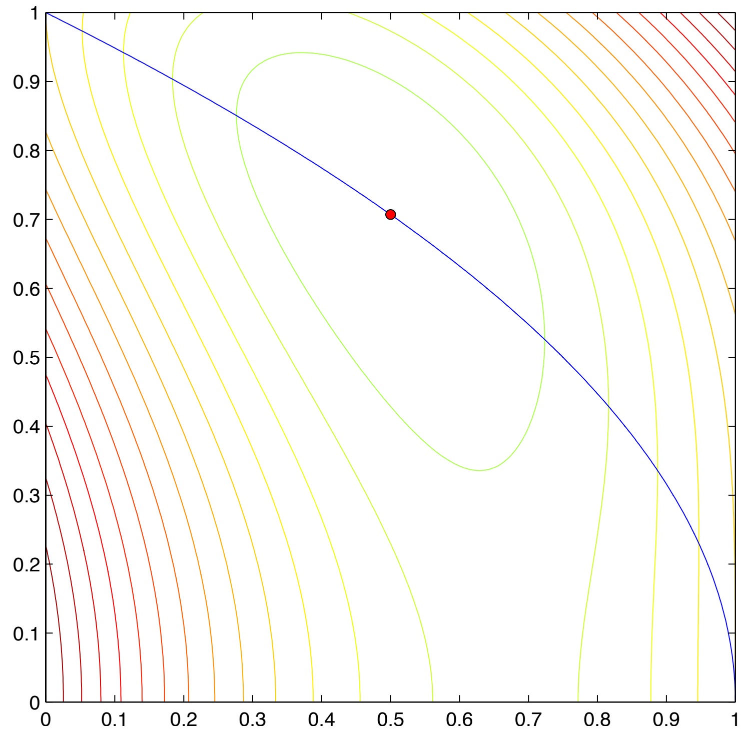
The augmented Lagrangian function for $\min x_1^2 + x_2^2$ subject to $x_1 + x_2^2 = 1$ for fixed $\sigma = 1$.

$\Phi(x, u, \sigma) = x_1^2 + x_2^2 - u(x_1 + x_2^2 - 1) + \frac{1}{2\sigma}(x_1 + x_2^2 - 1)^2.$

Contours of the augmented Lagrangian - an example...



$u = 0.99$



$u = y^* = 1$

The augmented Lagrangian function for $\min x_1^2 + x_2^2$ subject to $x_1 + x_2^2 = 1$ for fixed $\sigma = 1$.

$\Phi(x, u, \sigma) = x_1^2 + x_2^2 - u(x_1 + x_2^2 - 1) + \frac{1}{2\sigma}(x_1 + x_2^2 - 1)^2.$

Augmented Lagrangian methods

Th 22 $\Longrightarrow$ convergence guaranteed if u^k fixed and $\sigma^k \longrightarrow 0$

[similar to quadratic penalty methods]

$\Longrightarrow y^k \longrightarrow y^*$ and $c(x^k) \longrightarrow 0$

- check if $\|c(x^k)\| \leq \eta^k$ where $\eta^k \longrightarrow 0$
  - if so, set $u^{k+1} = y^k$ and $\sigma^{k+1} = \sigma^k$

    [recall expression of y^k in Th 22]
  - if not, set $u^{k+1} = u^k$ and $\sigma^{k+1} \leq \tau\sigma^k$ for some $\tau \in (0,1)$
- reasonable: $\eta^k = (\sigma^k)^{0.1+0.9j}$ where j iterations since σ^k last changed

Under such rules, can ensure that σ^k is eventually unchanged under modest assumptions, and (fast) linear convergence.

When σ^k is sufficiently large, need also to ensure that $\nabla^2\Phi(x^k, u^k, \sigma^k)$ is positive (semi-)definite.

A basic augmented Lagrangian method

Given $\sigma^0 > 0$ and u^0, let $k = 0$. Until "convergence" do:

- Set η^k and ϵ^{k+1}.
  If $\|c(x^k)\| \leq \eta^k$, set $u^{k+1} = y^k$ and $\sigma^{k+1} = \sigma^k$.
  Otherwise, set $u^{k+1} = u^k$ and $\sigma^{k+1} \leq \tau\sigma^k$.
- Starting from x_0^k (possibly, $x_0^k := x^k$), use an unconstrained minimization algorithm to find an "approximate" minimizer x^{k+1} of $\Phi(\cdot, u^{k+1}, \sigma^{k+1})$ for which $\|\nabla_x \Phi(x^{k+1}, u^{k+1}, \sigma^{k+1})\| \leq \epsilon^{k+1}$.
  Let $k := k + 1$. $\Diamond$

- Often choose $\tau = \min(0.1, \sqrt{\sigma^k})$
- Reasonable: $\epsilon^k = (\sigma^k)^{j+1}$, where j iterations since σ^k last changed