Augmented Lagrangian method

- Augmented Lagrangian
- Practical implementation
- Quadratic gradient projection

For $x \in \mathbb{R}^n$

$$\mathcal{L}_A = f(x) - \sum_{i\in\mathcal{E}} \lambda_i c_i + \frac{1}{2\mu} \sum_{i\in\mathcal{E}} c_i^2$$

Algorithmic framework For a given μ_k, λ^k, x_k, find x_{k+1} from approximate minimization

$$||\nabla \mathcal{L}_A(x_k, \mu_k)|| \le \tau_k,$$

Update λ_i^k by

$$\lambda_i^{k+1} = \lambda_i^k - c_i(x_k)/\mu_k$$

Choose new μ_{k+1} (optional, can keep old).
Example

$$\mathcal{L}_A = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2) + \frac{1}{2\mu}(x_1^2 + x_2^2 - 2)^2$$

Inequality constraints

$$\mathcal{L}_A = f(x) - \sum_{i\in\mathcal{I}} \lambda_i (c_i - s_i) + \frac{1}{2\mu} \sum_{i\in\mathcal{I}} (c_i - s_i)^2$$

with *slack variables* $s_i \ge 0$.
We use that (why?) $s_i = 0$ or $s_i = c_i - \mu\lambda_i^k$.
Theorem Let x^* be a local solution of the full problem with equality constraints (for simplicity). Suppose LICQ and second-order sufficient conditions are satisfied for $\lambda = \lambda^*$. Then there is a threshhold $\bar{\mu}$ such that for any $\mu \in (0, \bar{\mu}]$, x^* is a strict local minimizer of $\mathcal{L}_A$.
Proof Goal is to show

$$\nabla_x \mathcal{L}_A = 0, \text{ and } u^t \nabla_{xx}^2 \mathcal{L}_A u > 0$$

for $u \in \mathbb{R}^n$.
Since $c_i(x^*) = 0$, the first KKT condition is obvious: for any μ

$$\nabla f - \sum_{i\in\mathcal{A}(x^*)} \left[\lambda_i^* - \frac{c_i(x^*)}{\mu}\right] \nabla c_i(x^*)$$

$$= \nabla f - \sum_{i\in\mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*) = 0.$$

We have (why?)

$$\nabla_{xx}^2 \mathcal{L}_A = \nabla_{xx}^2 \mathcal{L} + \frac{1}{\mu} A^t A,$$

where a (full-rank)

$$A(x)^t = [\nabla c_1, \nabla c_2, \ldots, \nabla c_m].$$

Using $u = v + w$, $w \in \mathcal{N}A$, $v \in \mathcal{R}A^t$

$$u^t \nabla^2_{xx} \mathcal{L}_A u = +w^t \nabla^2_{xx} \mathcal{L} w + 2w^t \nabla^2_{xx} \mathcal{L} v + v^t \nabla^2_{xx} \mathcal{L} v + \frac{1}{\mu} ||Av||^2$$

Claim, if μ is small enough, then the above is always positive.
A more general statement is true: *Theorem* Instead of λ^*, x^* as in the previous theorem, we can take λ^k and x^k, sufficientely close to the former, so that the result of the previous theorem can be generalized.
Specifically:
· If

$$||\lambda^k - \lambda^*|| \leq \frac{\delta}{\mu_k}, \mu_k \leq \bar{\mu}$$

then there is (still) a unique minimizer x^k of $\mathcal{L}_A(x, \lambda^k, \mu_k)$ and

$$||x^k - x^*|| \leq M\mu_k ||\lambda^k_i - \lambda^*_i||$$

· If we update

$$\lambda^{k+1}_i = \lambda^k_i - c_i(x_k)/\mu_k,$$

then

$$||\lambda^{k+1}_i - \lambda^*_i|| \leq M\mu_k ||\lambda^k_i - \lambda^*_i||$$

· The matrix $\nabla^2 \mathcal{L}_A(x, \lambda^k, \mu_k)$ is positive definite and $\nabla c_i(x^k)$ are linearly independent.

Practical implementation LANCELOT

$$\min \mathcal{L}_A(x, \lambda, \mu), \text{ subject to } l \leq x \leq u.$$

First observation (see homework) that KKT becomes

$$P_{[l,u]} \nabla \mathcal{L}_A = 0$$

where

$$(P_{[l,u]})_i(g) = \begin{cases} \min(0, g_i) & x_i = l_i, \\ g_i & x_i \in (l_i, u_i), \\ \max(0, g_i) & x_i = u_i. \end{cases}$$

The algorithm Solve with (objective function) tolerance ω_k

$$||P_{[l,u]} \nabla \mathcal{L}_A|| \leq \omega_k$$

· If constraints are satisfied with (constraints) tolerance η_k

$$||c(x_k)|| \leq \eta_k$$

then *continue the search* of Lagrange multipliers:
update the Lagrange multipliers

$$\lambda^{k+1}_i = \lambda^k_i - c_i(x_k)/\mu_k,$$

don't change penalty

$$\mu_{k+1} = \mu_k,$$

tighten tolerances

$$\omega_{k+1} < \omega_k, \;\; \eta_{k+1} < \eta_k.$$

· If constraints are not satisfied with tolerance η_k

$$||c(x_k)|| > \eta_k$$

then *improve current* Lagrange multipliers:
don't change the Lagrange multipliers

$$\lambda_i^{k+1} = \lambda_i^k,$$

update penalty, by decreasing it

$$\mu_{k+1} < \mu_k,$$

tighten tolerances

$$\omega_{k+1} < \omega_k, \;\; \eta_{k+1} < \eta_k.$$

The only question left is how to solve the minimization of the objective function.

Approximate the augmented Lagrangian by a quadratic function and use the *quadratic gradient projection*

$$\min q(x) = \frac{1}{2} x^t G x + x^t d, \text{ subject to } l \le x \le u$$

The iteration is find the *Cauchy point* x^c, the optimal t_c so that $q(x^c)$, $x^c = x(t_c)$ is minimal for all

$$x(t) = P(x^0 + tg, l, u),$$

where $g = -\nabla q$, the gradient descent direction, and the projection

$$P(x,l,u)_i = \begin{cases} l_i, & x_i \le l_i, \\ x_i & x_i \in (l_i, u_i), \\ u_i & x_i \ge u_i. \end{cases}$$

After x^c is found $\rightarrow \mathcal{A}(x^c)$ (How?)
Use $\mathcal{A}(x^c)$ to to find an even better minimizer x^+ such that

$$x_i^+ = x_i^c, \; i \in \mathcal{A}(x^c), \text{ and } q(x^+) \le q(x^c)$$

In other words we solve

$$\min q(x) = \frac{1}{2} x^t G x + x^t d, \text{ subject to } x_i = x_i^c, \; i \in \mathcal{A}(x^c), l_i \le x_i \le u_i, i \notin \mathcal{A}(x^c).$$