Optimization Methods III
Augmented Lagrangian Methods and Interior-Point Methods

Anders Hansson and Martin Andersen

Linköping University and Technical Unversity of Denmark

March 26, 2024

Augmented Lagrangian Methods

Consider

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & h(x) = 0 \end{array} \tag{1}$$

where $f_0 : \mathbf{R}^n \to \mathbf{R}$, and where $h : \mathbf{R}^n \to \mathbf{R}^p$.

Equivalent problem

$$\begin{array}{ll} \text{minimize} & f_0(x) + \frac{\rho}{2}\|h(x)\|_2^2 \\ \text{subject to} & h(x) = 0 \end{array} \tag{2}$$

where $\rho \in \mathbf{R}_{++}$ called *penalty parameter*.

Augmented Lagrangian

Lagrangian for equivalent problem $L_\rho : \mathbf{R}^n \times \mathbf{R}^p \to \mathbf{R}$, called *augmented Lagrangian*

$$L_\rho(x, \mu) = f_0(x) + \mu^T h(x) + \frac{\rho}{2}\|h(x)\|_2^2$$

where $\rho \in \mathbf{R}_{++}$.

Dual function $g_\rho : \mathbf{R}^p \to \mathbf{R}$

$$g_\rho(\mu) = \inf_x L_\rho(x, \mu)$$

We will only discuss first order methods.

Dual Ascent

Use a gradient method to maximize g_ρ.

The solution to the original problem can be found from the optimal dual variable $\mu^\star$ by solving

$$\inf_x L_\rho(x, \mu^\star)$$

Key to dual ascent is to be able to compute a gradient or at least a sub-gradient of of g_ρ.

Sub-Gradient

Let $(x^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^p$ be current iterate and define next iterate $x^{k+1} \in \mathbf{R}^n$ as the solution of

$$\inf_x L_\rho(x, \mu^k)$$

Then

$$\begin{aligned}
g_\rho(\mu) &= \inf_x L_\rho(x, \mu) \\
&= \inf_x \left(f_0(x) + \mu^T h(x) + \frac{\rho}{2} \|h(x)\|_2^2 \right) \\
&\leq f_0(x^{k+1}) + \mu^T h(x^{k+1}) + \frac{\rho}{2} \|h(x^{k+1})\|_2^2 \\
&= f_0(x^{k+1}) + \left(\mu^k\right)^T h(x^{k+1}) + \left(\mu - \mu^k\right)^T h(x^{k+1}) + \frac{\rho}{2} \|h(x^{k+1})\| \\
&= g_\rho(\mu^k) + \left(\mu - \mu^k\right)^T h(x^{k+1})
\end{aligned}$$

and hence $h(x^{k+1})$ is a sub-gradient of g_ρ at μ^k.

Differentiable Case

If differentiability

$$0 = \nabla L_\rho(x^{k+1}, \mu^k) = \nabla f_0(x^{k+1}) + \nabla h(x^{k+1})^T \left(\mu^k + \rho h(x^{k+1})\right)$$

and natural update for μ is

$$\mu^{k+1} = \mu^k + \rho h(x^{k+1})$$

since this is dual ascent step which also makes μ^{k+1} dual feasible, i.e.

$$\nabla f_0(x^{k+1}) + \nabla h(x^{k+1})^T \mu^{k+1} = 0$$

Method of Multipliers

Given current iterate (x^k, μ^k) update as:

1. $x^{k+1} = \text{argmin}_x \, L_\rho(x, \mu^k)$
2. $\mu^{k+1} = \mu^k + \rho h(x^{k+1})$

It is possible to also update the parameter ρ:

$$\rho^{k+1} = \begin{cases} \beta \rho^k, & \|h(x^{k+1})\|_2 > \gamma \|h(x^k)\|_2 \\ \rho_k, & \|h(x^{k+1})\|_2 \leq \gamma \|h(x^k)\|_2 \end{cases}$$

where $\beta > 0$.

Constrained Nonlinear LS

Consider

$$\begin{array}{ll} \text{minimize} & \|F(x)\|_2^2 \\ \text{subject to} & G(x) = 0 \end{array}$$

where $F(x) = (f_1(x), \ldots, f_N(x))$ with $f_i : \mathbf{R}^n \to \mathbf{R}$, $i \in \mathbf{N}_N$ nonlinear residuals and $G(x) = (g_1(x), \ldots, g_N(x))$ with $g_i : \mathbf{R}^n \to \mathbf{R}$, $i \in \mathbf{N}_M$ nonlinear constraint functions.

Augmented Lagrangian:

$$\begin{aligned} L_\rho(x, \mu) &= \|F(x)\|_2^2 + \mu^T G(x) + \frac{\rho}{2}\|G(x)\|_2^2 \\ &= \|F(x)\|_2^2 + \frac{\rho}{2}\left\|G(x) + \frac{1}{\rho}\mu\right\|_2^2 - \frac{1}{2\rho}\|\mu\|_2^2 \\ &= \left\|\begin{bmatrix} F(x) \\ \sqrt{\rho/2}G(x) + \mu/\sqrt{2\rho} \end{bmatrix}\right\|_2^2 - \frac{1}{2\rho}\|\mu\|_2^2 \end{aligned}$$

Augmented Lagrangian Nonlinear LS problem

For fixed value of μ we can minimize L_ρ with respect to x by applying the Levenberg-Marquardt (LM)method to the unconstrained nonlinear LS problem

$$\text{minimize} \left\| \begin{bmatrix} F(x) \\ \sqrt{\rho/2}G(x) + \mu/\sqrt{2\rho} \end{bmatrix} \right\|_2^2$$

In practice we just run a few iterations of the LM method for each value of μ.

Linear Equality Constraints

Consider

$$\begin{array}{ll} \text{minimize} & F(x) + G(y) \\ \text{subject to} & Ax + By = c \end{array} \tag{3}$$

where $F : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^m \to \mathbf{R}$, $A \in \mathbf{R}^{p \times n}$, $N \in \mathbf{R}^{p \times m}$, and $c \in \mathbf{R}^p$.

We assume F and G convex but not necessarily differentiable.

Augmented Lagrangian $L_\rho : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$:

$$L_\rho(x, y, \mu) = F(x) + G(y) + \mu^T (Ax + By - c) + \frac{\rho}{2} \|Ax + By - c\|_2^2$$

Alternating Direction Methods of Mulptipliers (ADMM)

Block-coordinate minimization:

1. $x^{k+1} = \text{argmin}_x \, L_\rho(x, y^k, \mu^k)$
2. $y^{k+1} = \text{argmin}_y \, L_\rho(x^{k+1}, y, \mu^k)$
3. $\mu^{k+1} = \mu^k + \rho \left(Ax^{k+1} + By^{k+1} - c \right)$

Also ρ can be updated.

Often the minimizations in 1. and 2. are easier to carry out than the joint minimization in the method of multipliers.

Equivalent Formulation

Let $u^k = \mu^k/\rho$. Then equivalent scaled version of ADMM is

1. $x^{k+1} = \text{argmin}_x \left(F(x) + \frac{\rho}{2}\|Ax + By^k - c + u^k\|_2^2 \right)$
2. $y^{k+1} = \text{argmin}_y \left(G(y) + \frac{\rho}{2}\|Ax^{k+1} + By - c + u^k\|_2^2 \right)$
3. $u^{k+1} = u^k + Ax^{k+1} + By^{k+1} - c$

When $A = I$ we have

$$x^{k+1} = \mathbf{prox}_{\rho^{-1}F}\left(c - u^k - By^k\right)$$

A similar result holds for 2. when $B = I$.

Common Form of Optimization Problem

Consider

$$\text{minimize } F(x) + G(x)$$

with variable x, where $G(x) = \lambda \|x\|$ for some norm $\|\cdot\| : \mathbf{R}^m \to \mathbf{R}_+$ and where $\lambda \in \mathbf{R}_{++}$.

Equivalently

$$\begin{array}{ll} \text{minimize} & F(x) + G(y) \\ \text{subject to} & x - y = 0 \end{array} \tag{4}$$

with variables (x, y), where $y \in \mathbf{R}^n$.

ADMM updates

1. $x^{k+1} = \mathbf{prox}_{\rho^{-1}F}(y^k - u^k)$
2. $y^{k+1} = \mathbf{prox}_{\rho^{-1}G}(x^{k+1} + u^k)$
3. $u^{k+1} = u^k + x^{k+1} - y^{k+1}$

Proximal Operators

For 1-norm $f(x) = \sum_{i=1}^{m} |x_i|$ has

$$\mathbf{prox}_f(x)_i = \begin{cases} x_i - 1, & x_i \geq 1 \\ 0, & |x_i| \leq 1 \\ x_i + 1, & x_i \leq -1 \end{cases}$$

For nuclear norm of a matrix X, i.e. $f(X) = \|X\|_\star$

$$\mathbf{prox}_f(X) = \sum_{i=1}^{m} (\sigma_i - 1)_+ u_i v_i^T$$

where $X = \sum_{i=1}^{m} \sigma_i u_i v_i^T$ is an SVD of X and where $(\cdot)_+ : \mathbf{R} \to \mathbf{R}$ is the function defined by

$$(z)_+ = \begin{cases} z, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

Scaling Rule

If $f : \mathbf{R}^m \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ are functions related as $f(x) = \lambda g(x/\lambda)$, for $\lambda \in \mathbf{R}_{++}$, then

$$\mathbf{prox}_f(x) = \lambda\, \mathbf{prox}_{\lambda^{-1} g}(x/\lambda)$$

Sum of Functions

Assume $F_i : \mathbf{R}^n \to \mathbf{R}$ and

$$F(x) = \sum_{i=1}^{N} F_i(x)$$

Then equivalent optimization problem is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} F_i(x_i) + G(y) \\ \text{subject to} & x_i - y = 0, \qquad i \in \mathbf{N}_N \end{array} \tag{5}$$

with variables $(x_1, \ldots, x_N, y)$, where $x_i \in \mathbf{R}^n$

First update is

$$x_i^{k+1} = \mathbf{prox}_{\rho^{-1} F_i}(y^k - u_i^k), \qquad i \in \mathbf{N}_N$$

with $u = (u_1, \ldots, u_N)$.

Completing Squares

Second update is with variable y:

$$\text{minimize } G(y) + \frac{\rho}{2} \sum_{i=1}^{N} \left\| x_i^{k+1} - y + u_i^k \right\|_2^2$$

Expand the squared norm expressions, complete the squares with respect to y, and ignore terms independent of y:

$$\text{minimize } G(y) + \frac{N\rho}{2} \left\| y - \left(\bar{x}^{k+1} + \bar{u}^k \right) \right\|_2^2$$

where

$$\bar{x}^{k+1} = \frac{1}{N} \sum_{i=1}^{N} x_i^{k+1}; \qquad \bar{u}^k = \frac{1}{N} \sum_{i=1}^{N} u_i^k$$

Distributed ADMM

ADMM algorithm distributes:

1. $x_i^{k+1} = \mathbf{prox}_{\rho^{-1}F_i}(y^k - u_i^k), \qquad i \in \mathbf{N}_N$
2. $\bar{x}^{k+1} = \frac{1}{N}\sum_{i=1}^{N} x_i^{k+1}$
3. $\bar{u}^k = \frac{1}{N}\sum_{i=1}^{N} u_i^k$
4. $y^{k+1} = \mathbf{prox}_{(N\rho)^{-1}G}(\bar{x}^{k+1} + \bar{u}^k)$
5. $u_i^{k+1} = u_i^k + x_i^{k+1} - y^{k+1}, \qquad i \in \mathbf{N}_N$

Notice that 1. and 5. can be carried out in parallel.

General Constrained Optimization Problem

Consider

$$\begin{array}{ll} \text{minimize} & F(x) \\ \text{subject to} & x \in C \end{array} \tag{6}$$

where $F : \mathbf{R}^n \to \mathbf{R}$ convex function for which proximal operator can be evaluated cheaply, and where C is convex set.

Equivalent problem

$$\begin{array}{ll} \text{minimize} & F(x) + I_C(y) \\ \text{subject to} & x - y = 0. \end{array} \tag{7}$$

The prox operator of the indicator function of a convex set obtained from

$$\text{minimize}_{y \in C} \|y - x\|_2^2$$

which is the projection of x onto C.

Barrier Problem

Consider

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i \in \mathbf{N}_m \\ & Ax = b \end{array} \tag{8}$$

where $f_0 : \mathbf{R}^n \to \mathbf{R}$, $f_i : \mathbf{R}^n \to \mathbf{R}$, $i \in \mathbf{N}_m$ are convex functions, and where $A \in \mathbf{R}^{p \times n}$ and $b \in \mathbf{R}^p$.

Replace inequality constraints with *barrier function* $\phi : \mathbf{R}^n \to \mathbf{R}$ where

$$\phi(x) = -\sum_{i=1}^{m} \ln\left(-f_i(x)\right)$$

Barrier problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) + \frac{1}{t}\phi(x) \\ \text{subject to} & Ax = b \end{array} \tag{9}$$

where $t > 0$.

Optimality Conditions

We have

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x)$$

$$\nabla^2 \phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

Lagrangian $L : \mathbf{R}^n \times \mathbf{R}^p$ of (9) defined by

$$L(x, \lambda) = f_0(x) + \frac{1}{t}\phi(x) + \mu^T (Ax - b)$$

Hence optimality conditions for (9) are

$$\nabla f_0(x) - \sum_{i=1}^{m} \frac{1}{t f_i(x)} \nabla f_i(x) + A^T \mu = 0$$

$$Ax = b$$

Interior-Point (IP) Method

Conditions can be solved with Newton's method by linearization.

Equations for search directions $(\Delta x, \Delta\mu) \in \mathbf{R}^x \times \mathbf{R}^p$:

$$\left(\nabla^2 f_0(x) + \frac{1}{t}\nabla^2\phi(x)\right)\Delta x + A^T\Delta\mu = -\nabla f_0(x) - \frac{1}{t}\nabla\phi(x) - A^T\mu$$

$$A\Delta x = -(Ax - b)$$

where (x, μ) is the current iterate.

Use Newton's method to solve the barrier problem, increase the value of t and repeat.

Termination Criterion

Define the Lagrangian $\mathcal{L} : \mathbf{R}^n \times^m \times \mathbf{R}^p \to \mathbf{R}$ for (8) as

$$\mathcal{L}(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \mu^T (Ax - b)$$

If $x^\star$ solves barrier problem, then by optimality conditions above for (9) there exist μ:

$$\nabla f_0(x^\star) - \sum_{i=1}^{m} \frac{1}{t f_i(x^\star)} \nabla f_i(x^\star) + A^T \mu = 0$$

and hence $x^\star$ minimizes $\mathcal{L}(x, \lambda^\star, \mu)$ where $\lambda_i^\star = 1/\left(-t f_i(x^\star)\right) \geq 0$.

Let $g : \mathbf{R}^m \times \mathbf{R}^p$ be Lagrange dual function, i.e. $g(\lambda, \mu) = \min_x \mathcal{L}(x, \lambda, \mu)$. By weak duality

$$p^\star \geq g(\lambda^\star, \mu) = \mathcal{L}(x^\star, \lambda^\star, \mu) = f_0(x^\star) - m/t$$

Above $p^\star$ is the optimal value of (8). Hence $x^\star$ from the barrier problem is ϵ-suboptimal with $\epsilon = m/t$.

Strictly Feasible Initial Point

Phase I problem:

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, \quad i \in \mathbf{N}_m \\ & Ax = b \end{array}$$

with variables $(x, s) \in \mathbf{R}^{n+1}$.

If optimal solution of this problem has $s^\star < 0$, then corresponding $x^\star$ is strictly feasible for (8).

Other IP Methods

Efficient IP methods update the primal and dual variables jointly and are called *primal-dual* IP methods.

Specifically the corresponding barrier problem is not solved to optimality before t is updated.

Barrier Problem for $\mathcal{S}^n_{++}$

Consider

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}\, CX - \ln\det X \\ \text{subject to} & \mathbf{tr}\, A_i X = b_i, \qquad i \in \mathbf{N}_p \end{array} \tag{10}$$

where $X \in \mathbf{S}^n_{++}$, $A_i \in \mathbf{S}^n$, $b_i \in \mathbf{R}$, and $C \in \mathbf{S}^n$.

Lagrangian $L : \mathbf{S}^n_+ \times \mathbf{R}^p$ given by

$$L(X, y) = \mathbf{tr}\, CX - \ln\det X + \sum_{i=1}^{p} y_i \left(\mathbf{tr}\, A_i X - b_i\right)$$

Optimality conditions

$$C - X^{-1} + \sum_{i=1}^{p} y_i A_i = 0$$

$$\mathbf{tr}\, A_i X - b_i = 0, \qquad i \in \mathbf{N}_p$$

Equivalent Otimality Conditions

Let $X = E(x) = \sum_{i=1}^{m} x_i E_i$ where E_i are basis matrices for a symmetric $n \times n$ matrix, $m = n(n+1)/2$ and where $E : \mathbf{R}^m \to \mathbf{S}^n$ and $x = (x_1, \ldots, x_m)$.

Equivalent optimality conditions

$$\begin{aligned} \mathbf{tr}\, CE_k - \mathbf{tr}\, E(x)^{-1} E_k + \sum_{i=1}^{p} y_i\, \mathbf{tr}\, A_i E_k &= 0, \qquad k \in \mathbf{N}_m \\ \mathbf{tr}\, A_i X - b_i &= 0, \qquad i \in \mathbf{N}_p \end{aligned}$$

Newton's Method

Conditions can be solved with Newton's method by linearization.

Equations for search directions $(\Delta x, \Delta y) \in \mathbf{R}^m \times \mathbf{R}^p$:

$$\sum_{j=1}^{m} \mathbf{tr}\, E_k X^{-1} E_j X^{-1} \Delta x_j + \sum_{i=1}^{p} \mathbf{tr}\, A_i E_k \Delta y_i = -\mathbf{tr}\, C E_k + \mathbf{tr}\, X^{-1} E_k - \sum_{i=1}^{p} y_i\, \mathbf{tr}\, A_i E_k, \qquad k \in \mathbf{N}_m$$

$$\sum_{j=1}^{m} \mathbf{tr}\, A_i E_j \Delta x_j = -\mathbf{tr}\, A_i X + b_i, \qquad i \in \mathbf{N}_p$$

where (X, y) is the current iterate.

Cost of solving equations minor compared to forming the equations.

Conic Optimization Problem

Consider

$$\begin{array}{lll} \text{minimize} & \mathbf{tr}\, CX & \\ \text{subject to} & \mathbf{tr}\, A_i X = b_i, & i \in \mathbf{N}_p \\ & X \in \mathbf{S}_+^n & \end{array} \tag{11}$$

Lagrangian $\mathcal{L} : \mathbf{S}^n \times \mathbf{S}^n \times \mathbf{R}^p$ is

$$\mathcal{L}(X, Z, y) = \mathbf{tr}\, CX - \mathbf{tr}\, ZX + \sum_{i=1}^{p} y_i \left(\mathbf{tr}\, A_i X - b_i\right)$$

Lagrange Dual Problem

We have

$$\mathcal{L}(X, Z, y) = \mathbf{tr}\left(C + \sum_{i=1}^{p} y_i A_i - Z\right) X - b^T y$$

where $y = (y_1, \dots, y_p)$ and $b = (b_1, \dots, b_p)$.

Lagrange dual problem

$$\begin{array}{ll} \text{maximize} & -b^T y \\ \text{subject to} & C + \sum_{i=1}^{p} y_i A_i = Z \\ & Z \in \mathbf{S}_+^n \end{array} \tag{12}$$

Barrier Function

Let $X^\star$ be optimal for (10). Then $X^\star$ primal feasible and there $\exists\, y$:

$$C - (X^\star)^{-1} + \sum_{i=1}^{p} y_i A_i = 0$$

which is first row of optimality conditions for (10).

Moreover $Z = (X^\star)^{-1}$ is dual feasible since $X^\star \in \mathbf{S}^n_{++}$. Hence by the optimality conditions for (10) the duality gap is

$$\begin{aligned}\mathbf{tr}\, CX^\star + b^T y &= \mathbf{tr}\left((X^\star)^{-1} - \sum_{i=1}^{p} y_i A_i\right) X^\star + b^T y \\ &= n - \sum_{i=1}^{p} y_i\, \mathbf{tr}\, A_i X^\star + b^T y = n\end{aligned}$$

If $\ln\det X$ in (10) replaced with $\frac{1}{t}\ln\det X$ duality gap is n/t, showing that $\ln\det X$ barrier function.

Logarithmic Barrier Function

Function $\psi : K \to \mathbf{R}$, called *generalized logarithm* for a proper cone K if

1. $\mathbf{dom}\,\psi = \mathbf{int} K$
2. $\nabla\psi(x) \in K^* \ \forall\ x \in K$
3. $\exists\ \theta$ such that $\psi(sx) = \psi(x) + \theta \ln s \ \forall\ x \in K$ and $s > 0$.

Here θ is called the degree of ψ.

For any generalized logarithm it holds that for $x \in K$ we have $\langle x, \nabla\psi(x)\rangle = \theta$. The duality gap will be θ/t for solutions from the barrier problem.

Generalized logiarithms for different cones

- $\psi(x) = \sum_{i=1}^{n} \ln(x_i)$, where $x = (x_1, \ldots, x_n) \in \mathbf{R}_+^n$ with $\theta = n$.
- $\psi(X) = \ln\det X$, where $X \in \mathbf{S}_+^n$ with $\theta = n$.
- $\psi(x) = \ln\left(y_n^2 - y_1^2 - \cdots - y_{n-1}^2\right)$, where $x \in \mathbf{Q}^n$ with $\theta = 2$.
- $\psi(x) = \ln\left(y\ln(z/y) - x\right) + \ln y + \ln z$, where $(x, y, z) \in K_{\exp}$ with $\theta = 3$.