Lecture 28: Alternating Direction Method of Multipliers (ADMM)

- Well-suited to $\begin{cases} \text{distributed optimization} \\ \text{large-scale problems} \end{cases}$

- Precursors
  - ⋆ Dual ascent
  - ⋆ Dual decomposition
  - ⋆ Method of multipliers

- Design of optimal sparse feedback gains via ADMM

  Lin, Fardad, Jovanović, IEEE TAC '11 (submitted; also: arXiv:1111.6188v1)

- Online resources
  - ⋆ Stephen Boyd's webpage

    ADMM material (paper, talks, Matlab files)

    ℓ_1 methods for convex-cardinality problems (lectures and videos)

Equality-constrained convex optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & A\,x \;=\; b \end{array}$$

$f\colon \mathbb{R}^n \to \mathbb{R}$ – convex function

- Lagrangian

$$\mathcal{L}(x,y) \;=\; f(x) \,+\, y^T\,(A\,x \,-\, b)$$

- dual function

$$g(y) \;=\; \inf_x \, \mathcal{L}(x,y)$$

- dual problem

$$\text{maximize} \;\; g(y)$$

Dual ascent

$$
\begin{array}{rcl}
x\text{-minimization:} & x^{k+1} & := \; \underset{x}{\arg\min} \; \mathcal{L}(x, y^k) \\
\text{dual variable update:} & y^{k+1} & := \; y^k \; + \; s^k \left(A\, x^{k+1} \; - \; b\right)
\end{array}
$$

- Features
  - ⋆ For properly selected $s^k \;\Rightarrow\; g(y^{k+1}) \;>\; g(y^k)$
  - ⋆ Requires strong assumptions
  - ⋆ May converge slowly
  - ⋆ Can lead to distributed implementation

Dual decomposition

separable form:

$$f(x) = \sum_{n=1}^{N} f_n(x_n)$$

Lagrangian:

$$\begin{aligned} \mathcal{L}(x,y) &= \sum_{n=1}^{N} f_n(x_n) + y^T \left(\sum_{n=1}^{N} A_n x_n - b \right) \\ &= \sum_{n=1}^{N} \mathcal{L}_n(x_n, y) - y^T b \end{aligned}$$

decomposition:

$$\mathcal{L}_n(x_n, y) := f_n(x_n) + y^T A_n x_n$$

- **Can be solved in parallel**

DUAL DECOMPOSITION:

$$
\begin{aligned}
x_n^{k+1} &:= \underset{x_n}{\arg\min} \; \mathcal{L}_n(x_n, y^k) \\
y^{k+1} &:= y^k \; + \; s^k \left(\sum_{n\,=\,1}^{N} A_n x_n^{k+1} \; - \; b \right)
\end{aligned}
$$

- distributed optimization
  - ⋆ broadcast y^k
  - ⋆ update x_n^{k+1} in parallel
  - ⋆ gather $A_n x_n^{k+1}$
- well-suited to large-scale problems
  - ⋆ sub-problems solved iteratively in parallel
  - ⋆ dual variable update provides coordination

Method of multipliers

augmented Lagrangian: $\mathcal{L}_\rho(x,y) \;=\; \mathcal{L}(x,y) \;+\; \frac{\rho}{2}\,\|Ax \;-\; b\|_2^2$

METHOD OF MULTIPLIERS:

$$\begin{aligned} x^{k+1} &:= \operatorname*{arg\,min}_{x} \; \mathcal{L}_\rho(x, y^k) \\ y^{k+1} &:= y^k \;+\; \rho\,\big(Ax^{k+1} \;-\; b\big) \end{aligned}$$

compared to dual ascent:

- advantages:
  - ⋆ convergence under milder assumptions
  - ⋆ brings robustness
- disadvantage
  - ⋆ quadratic term: in general not separable $\Rightarrow$ may not be solved in parallel

OPTIMALITY CONDITIONS:

$$
\begin{aligned}
\nabla_x \mathcal{L}_\rho(x^\star, y^\star) &= \nabla_x f(x^\star) \,+\, A^T y^\star \,=\, 0 \\
\nabla_y \mathcal{L}_\rho(x^\star, y^\star) &= A x^\star \,-\, b \,=\, 0
\end{aligned}
$$

- x^{k+1} minimizer of $\mathcal{L}(x, y^k)$

$$
\begin{aligned}
0 &= \nabla_x \mathcal{L}(x^{k+1}, y^k) \\
&= \nabla_x f(x^{k+1}) \,+\, A^T y^k \,+\, \rho A^T (A x^{k+1} - b) \\
&= \nabla_x f(x^{k+1}) \,+\, A^T (y^k \,+\, \rho (A x^{k+1} - b)) \\
&= \nabla_x f(x^{k+1}) \,+\, A^T y^{k+1}
\end{aligned}
$$

- dual feasibility satisfied at every iteration
- primal feasibility satisfied in the limit

$$
\lim_{k \to \infty} A\, x^k \,=\, b
$$

Alternating direction method of multipliers

- Converges under mild assumptions
  - ⋆ robust dual decomposition
- Facilitates decomposition
  - ⋆ decomposable method of multipliers
- Proposed in ’70s
- Many modern applications
  - ⋆ distributed computing
  - ⋆ distributed signal processing
  - ⋆ image denoising
  - ⋆ machine learning
  - ⋆ statistics

Boyd et al., Foundations and Trends in Machine Learning ’11

- standard ADMM formulation

$$\begin{array}{ll} \text{minimize} & f(x) \; + \; g(z) \\ \text{subject to} & Ax \; + \; Bz \; = \; c \end{array}$$

augmented Lagrangian

$$\mathcal{L}_\rho(x,z,y) \; = \; f(x) \; + \; g(z) \; + \; y^T(Ax \; + \; Bz \; - \; c) \; + \; \frac{\rho}{2} \|Ax \; + \; Bz \; - \; c\|_2^2$$

ADMM:

$$\begin{array}{rcl} x^{k+1} & := & \arg\min_x \mathcal{L}_\rho(x, z^k, y^k) \\ z^{k+1} & := & \arg\min_z \mathcal{L}_\rho(x^{k+1}, z, y^k) \\ y^{k+1} & := & y^k \; + \; \rho\left(Ax^{k+1} \; + \; Bz^{k+1} \; - \; c\right) \end{array}$$

Reduces to method of multipliers if minimization done jointly (over x and z)

OPTIMALITY CONDITIONS:

$$\begin{array}{rcl} \nabla_x \mathcal{L}_\rho(x^\star, y^\star, z^\star) & = & \nabla_x f(x^\star) \; + \; A^T y^\star \; = \; 0 \\ \nabla_z \mathcal{L}_\rho(x^\star, y^\star, z^\star) & = & \nabla_z g(z^\star) \; + \; B^T y^\star \; = \; 0 \\ \nabla_y \mathcal{L}_\rho(x^\star, y^\star, z^\star) & = & A x^\star \; + \; B z^\star \; - \; c \; = \; 0 \end{array}$$

- z^{k+1} minimizes $\mathcal{L}(x^{k+1}, z, y^k)$

$$\begin{array}{rcl} 0 & = & \nabla_z g(z^{k+1}) \; + \; B^T y^k \; + \; \rho B^T (A x^{k+1} \; + \; B z^{k+1} \; - \; c) \\ & = & \nabla_z g(x^{k+1}) \; + \; B^T y^{k+1} \end{array}$$

- second dual feasibility satisfied at every iteration
- primal and first dual feasibility satisfied asymptotically