Augmented Lagrangian Methods (ALMs) for Mixed-Integer Linear Programming (MILP)

Wotao Yin

Decision Intelligence Lab, Alibaba DAMO Academy

April 10, 2020

Goals

Solve arbitrary-scale structured MILPs:

- Distributed decision making under shared resource constraints
- Future decision making under random factors
- Optimization on networks

Generalize popular ALMs and ADMMs to MILPs:

- Decompose MILPs to simpler subproblems
- Re-engineer the methods to ensure convergence
- Solve subproblems by existing techniques

Outline

- Dual decomposition, Primal decomposition
- Augmented Lagrangian method (ALM)
- Alternating direction method of multipliers (ADMM)
- Challenges due to mixed integers
- Exact-penalty augmented Lagrangian
- ADMM for MILP

Dual decomposition

- Primal problem

$$\begin{aligned} \underset{x_i}{\text{minimize}} \quad & \underline{\sum_{i=1}^n f_i(x_i)} \\ \text{subject to} \quad & \underline{\sum_{i=1}^n A_i x_i = b} \end{aligned}$$

$x_1 \cdots x_n$

u

- Lagrangian: $L(x_1, \ldots, x_n, \underline{u}) = \sum_{i=1}^n (f_i(x_i) + \langle \underline{u}, A_i x_i \rangle) - \langle u, b \rangle$
- Add $\underline{u \geq 0}$ if constraints are $\sum_{i=1}^n A_i x_i \leq b$
- Dual decomposition method:

$$\begin{aligned} \underline{x_i^{k+1}} &\in \arg\min f_i(x_i) + \langle \underline{u^k}, A_i x_i \rangle, \quad \underline{i = 1, \ldots, n} \\ u^{k+1} &= u^k + \alpha \underline{\left(\sum_{i=1}^n A_i x_i - b \right)} \end{aligned}$$

- If constraints are $\underline{\sum_{i=1}^n A_i x_i \leq b}$, then change the 2nd line to

$$u^{k+1} = \max \left\{ u^k + \alpha \left(\sum_{i=1}^n A_i x_i - b \right), 0 \right\}$$

Pros:

- Simpler subproblems
- Parallel subproblems

Cons:

- Slow convergence, even if f_i's are convex
- Requires conditions for x^k to converge

Primal decomposition (not used in later slides)

- Same primal problem

$$\begin{aligned} \underset{x_i}{\text{minimize}} \quad & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} \underbrace{A_i x_i}_{y_i} = b \end{aligned}$$

← diff to proj onto

- Introduce $\phi_i(y_i) = \inf_{x_i} \{ f_i(x_i) : A_i x_i = \cancel{b}\, y_i \}$
- Equivalent problem

$p_i \in \partial \phi_i(y_i)$ parallel

$$\begin{aligned} \underset{y_1,\dots,y_n}{\text{minimize}} \quad & \sum_{i=1}^{n} \phi_i(y_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} y_i = b \end{aligned}$$

← easy to proj

The objective functions are independent, and the constraint is easy to project onto. Apply the subgradient-projection method.

E. Ryu. W. Yin

Augmented Lagrangian

- Primal problem
$$\begin{array}{ll} \underset{x_i}{\text{minimize}} & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to} & \sum_{i=1}^{n} A_i x_i = b \end{array}$$

x_i integer

- Write $Ax = \sum_{i=1}^{n} A_i x_i$
- Add $\frac{\beta}{2}\|Ax - b\|^2$ to L to get *augmented Lagrangian*

$$L^{\mathsf{A}}(x_1, \ldots, x_n, u) = L(x_1, \ldots, x_n, u) + \frac{\beta}{2}\|Ax - b\|^2$$

- Squared 2-norm is typical, but 1-norm and ∞-norm also used.

Observe:

- $\underset{x_1 \cdots x_n}{\arg\min}\, L(x_1, \ldots, x_n, u^k)$ decomposes to independent $\arg\min_{x_i}$
- $\underset{x_1 \cdots x_n}{\arg\min}\, L^{\mathsf{A}}(x_1, \ldots, x_n, u^k)$ does not decompose

So, why bother with augmented Lagrangian?

- Define dual functions, which are concave:

$$d(u) = \inf_{x_1,\ldots,x_n} L(x_1,\ldots,x_n,u),$$
$$d^{\mathsf{A}}(u) = \inf_{x_1,\ldots,x_n} L^{\mathsf{A}}(x_1,\ldots,x_n,u)$$

if x is feas

i.e. $Ax - b = 0$

$\langle u, Ax - b \rangle = 0$

$\|Ax - b\| = 0$

- For every u, L and L^{A} are constraint relaxations
- For every u, $d(u)$ and $d^{\mathsf{A}}(u)$ are lower bounds of the original problem.
- Dual problem looks for the highest lower bound

$$\operatorname*{maximize}_u d(u)$$

Hope $d(u^*)$ is tight

$=$ opt. original obj

Compare dual-decomposition method (DDM):

$$x_i^{k+1} \in \arg\min f_i(x_i) + \langle u^k, A_i x_i - y \rangle, \quad i = 1, \dots, n$$

$$u^{k+1} = u^k + \underset{=}{\alpha}(\underbrace{\sum_{i=1}^{n} A_i x_i^{k+1} - y}_{\in \partial d(u^k)})$$

with augmented Lagrangian method (ALM):

$$x^{k+1} \in \arg\min_x L^{A}(x_1, \dots, x_n, u^k)$$

$$u^{k+1} = u^k + \beta(\sum_{i=1}^{n} A_i x_i - y)$$

If $f_1, \dots, f_n$ are convex, closed, and proper, then,

- DDM is equivalent to "subgradient ascent" (slow, unstable) for $d(u)$:

$$u^{k+1} = u^k + \underset{=}{\alpha} p, \quad p \in \partial d(u^k)$$

concave

- ALM is equivalent to proximal ascent (fast, stable) for $d(u)$:

$$u^{k+1} = \arg\min_u d(u) + \frac{1}{2\beta}\|u - u^k\|^2$$

What if nonconvex?

Perturbing the RHS

min $f(x)$
s.t. $Ax = b + \delta b$

- Define

$$p(\Delta_b) = \inf_x \{f(x) : Ax = b + \Delta_b\}$$

- $p(\Delta_b)$ is convex if f is convex; it is nonconvex in general.
- The duality gap $p(0) - \sup_u d(u)$ is determined by the geometry of $p(\Delta_b)$:

orig opt obj

- Recall dual function $d(u) = \inf_x L(x, u) = \inf_x f(x) + \langle u, Ax - b \rangle$ $-\delta b$
- $p(\Delta_b) \geq d(u) - \langle u, \Delta_b \rangle$ for any Δ_b
- $d(u) = \inf_{x,\Delta_b} \{f(x) + \delta_{\Delta_b = Ax - b} + \langle u, \Delta_b \rangle\} = \inf_{\Delta_b} \{p(\Delta_b) + \langle u, \Delta_b \rangle\}$

$\Rightarrow$ Affine function $\ell(\Delta_b) = d(u) - \langle u, \Delta_b \rangle$ is the tight lower bound of $p(\Delta_b)$ that obeys $\ell(0) = d(u)$

Geometric view: convex case

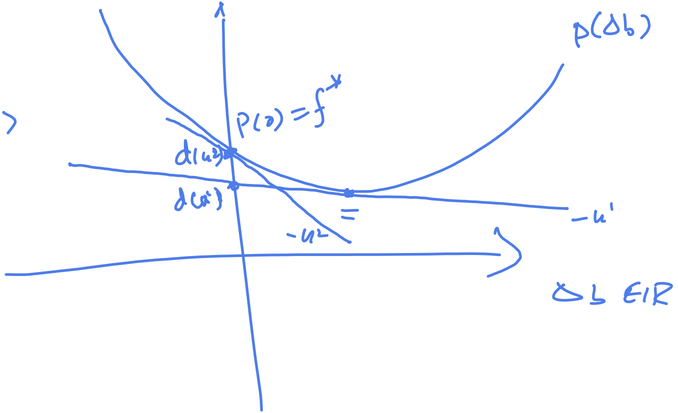
f convex

$d(u^2) = p(0) = f^*$

fix $u = u'$

$d(u') - \langle u', \Delta b \rangle$

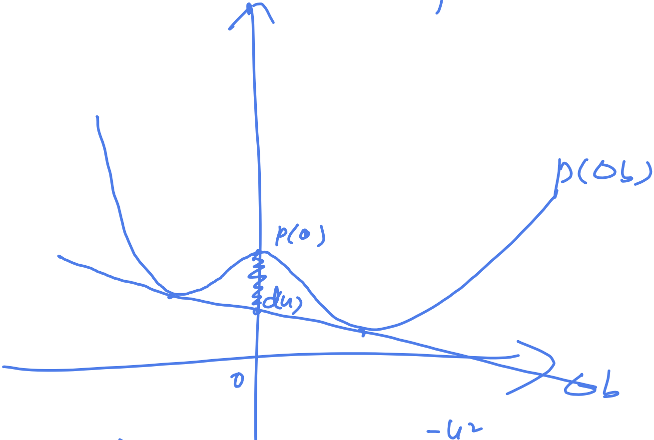
$u = u^2$



by change u in $d(u)$, attain $d(u) = p(0)$

Geometric view: nonconvex case

f noncvx



$d(u^2) = p(0)$

no sup

$d(u) < p(0)$

ALM narrows duality gap

cvx

concave

- Recall

$$p(\Delta_b) = \inf_x \{f(x) : Ax = b + \Delta_b\}$$

- Recall dual function of augmented Lagrangian

$$\begin{aligned} d^{\mathsf{A}}(u) &= \inf_x L^{\mathsf{A}}(x, u) \\ &= \inf_x f(x) + \langle u, Ax - b\rangle + \frac{\beta}{2}\|Ax - b\|^2 \end{aligned}$$

- $p(\Delta_b) \geq d^{\mathsf{A}}(u) - \langle u, \Delta_b\rangle - \frac{\beta}{2}\|\Delta_b\|^2$ for any Δ_b
- $d(u) = \inf_{\Delta_b}\{p(\Delta_b) + \langle u, \Delta_b\rangle + \frac{\beta}{2}\|\Delta_b\|^2\}$

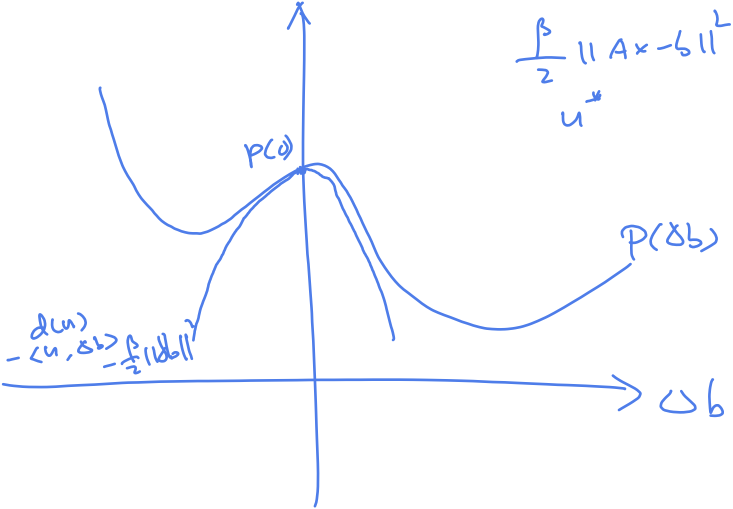
$\ell(\Delta_b) = d(u) - \langle u, \Delta_b\rangle$

$\Rightarrow$ Concave quadratic function $\ell^{\mathsf{A}}(\Delta_b) = d^{\mathsf{A}}(u) - \langle u, \Delta_b\rangle - \frac{\beta}{2}\|\Delta_b\|^2$ is the tight lower bound of $p(\Delta_b)$ that obeys $\ell^{\mathsf{A}}(0) = d(u)$

Geometric view: nonconvex case

$\frac{\beta}{2}\|Ax-b\|^2$

u^*



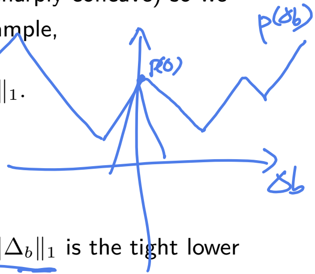
Mixed-integer linear constraint

- Define bounded $M \subset \mathbb{R}^{n_1} \times \mathbb{Z}^{n_2}$, assume rational data, and re-define

$$p(\Delta_b) = \inf_x \{c^\top x : Ax = b + \Delta_b,\ x \in M\}$$

At $\Delta_b = 0$, $p(\Delta_b)$ can be sharp and nonconvex (e.g., sharply concave) so we need a sharp penalty in augmented Lagrangian, for example,

$$L^1(x, u) = L(x, u) + \frac{\beta}{2}\|Ax - b\|_1.$$



(∞-norm also works)

- Define $d^1(u) = \inf_{x \in M} L^1(x, u)$.

$\Rightarrow$ Polyhedral function $\ell^1(\Delta_b) = d^1(u) - \langle u, \Delta_b \rangle - \frac{\beta}{2}\|\Delta_b\|_1$ is the tight lower bound of $p(\Delta_b)$ that obeys $\ell^1(0) = d(u)$

$\Rightarrow$ For sufficiently large β, duality gap $p(0) - \sup_u d(u)$ is zero (under technical regularity conditions).

Geometric view: concave polyhedral case

Exact penalty

Theorem (Boland-Eberhard'15)

Let ϕ be a convex and monotonically increasing function. $\mathbb{R}_+ \to \mathbb{R}_+$
Under mild technical assumptions, augmented Lagrangian with penalty function $\phi(\|Ax - b\|)$ has asymptotically 0 duality gap, that is,

$$\lim_{\beta \to \infty} \sup_u d^{\beta,\phi(\|\cdot\|)}(u) = p(0).$$

Theorem (Feizollahi-Ahmed-Sun'17)

For a finite set M, there exists a finite β to attain 0 duality gap, that is,

$$\sup_u d^{\beta,\cancel{\phi}(\|\cdot\|)}(u) = p(0).$$

ADMM

ALM

no duality

- Recall ALM cannot decompose into independent problems.

$f_1(x_1) = f_{1,1}(x_{1,1}) \cdots f_{1,p}(x_{1,p})$ A_1

- Consider $n = 2$.
- Primal problem

$$\begin{aligned} \underset{x_1, x_2}{\text{minimize}} \quad & f_1(x_1) + f_2(x_2) \\ \text{subject to} \quad & A_1 x_1 + A_2 x_2 = b \end{aligned}$$

- ADMM iteration

$f_1(x_1) + f_2(x_2) + \langle u, A_1x_1 + A_2x_2 - b \rangle + \frac{\beta}{2}\|A_1x_1 + A_2x_2 - b\|_2^2$

$$\begin{aligned} x_1^{k+1} &\in \arg\min L^{\mathsf{A}}(x_1, x_2^k, u^k) \\ x_2^{k+1} &\in \arg\min L^{\mathsf{A}}(x_1^{k+1}, x_2, u^k) \\ u^{k+1} &= u^k + \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{aligned}$$

- Converges under convexity and technical regularity conditions.
- Convergence speed is comparable to ALM.
- With proper modifications, extend to $n \geq 3$, e.g., "Jacobi ADMM"
- How about nonconvexity and mixed-integer set?

Nonconvex ADMM

Moreau ALM

Wang, **Y.**, Zeng "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization" for sufficient conditions:

- There are nonconvex examples on which ALM and ADMM both diverge
- It may appear ALM is "easier"" to converge than ADMM. But, not true.
- There exists examples: ADMM converges but ALM diverges (with fixed β)
- Key enablers: for noncvx f
  - $\mathrm{Im}(A_1) \subseteq \mathrm{Im}(A_2)$
  - f_2 is Lipschitz differentiable
  - $f_1(x_1) + f_2(x_2)$ is coercive over $\{(x_1, x_2) : A_1x_1 + A_2x_2 = b\}$

$\|(x_1,x_2)\| \to \infty$

$\Rightarrow f_1(x_1) + f_2(x_2) \to \infty$

Mixed-integer ADMM

Consider 2-block **MILP**

$$\begin{aligned} p^* = \min_{x,z} \quad & c^\top x + g^\top z \\ \text{s.t.} \quad & Ax + Bz = 0 \\ & x \in X, z \in Z \end{aligned}$$

where

- constraints X and Z are bounded and may be mixed-integer sets
- the problem is feasible
- c, g, A, B, X, Z are rational

- 1-norm augmented Lagrangian

$$L^1(x, z, u, \beta) = c^\top x + g^\top z + \langle u, Ax + Bz \rangle + \frac{\beta}{2} \|Ax + Bz\|_1$$

Iterate:

Step 1: let $x^{k+1} \in \arg\min_{x \in X} L^1(x, z^k, u^k, \beta^k)$; $\Rightarrow$ many parallel subprobs

Step 2: add an `ALCut`$(x^{k+1}, z^k, u^k, \beta^k)$ to z-subproblem and solve for z^{k+1};

Step 3: update u^{k+1}, β^{k+1}.

$z \in Z$

- Both x-subproblems and z-subproblems are MILPs.
- *AL cuts* are inequalities added to the z-subproblems, so their objectives approximate $\min_{x \in X} L(x, z, \mu^k, \beta^k)$ with increasing accuracy.

- Collect x terms in L^1 and minimize over x

$$\min_{z\in Z} \quad R_\rho(z) = \min_{x\in X} c^\top x + \rho\|Ax+Bz\|_1 \quad = p^*$$

- Define

$$P(z,u,\beta) = \underbrace{\min_{x\in C} L^1(x,z,u,\beta) - g^\top z}_{\text{half dual func.}}$$

- Suppose $\|u\|_\infty + \beta \leq \rho$

$\Rightarrow$

- Weak duality $P(z,u,\beta) \leq R_\rho(z)$
- Strong duality $\max_{\|u\|_\infty+\beta\leq\rho} P(z,u,\beta) = R_\rho(z)$
- Given $\bar{z}\in Z$, $\bar{u}$, and $\bar{\beta}$, function $R_\rho(z)$ has a lower bound (the AL cut)

$$R_\rho(z) \geq \underbrace{P(\bar{z},\bar{\mu},\bar{\beta}) + \langle\bar{\mu}, Bz - B\bar{z}\rangle - \bar{\beta}\|Bz - B\bar{z}\|_1}_{l(z)}.$$

MILP ADMM

Iterate:

for β^k large eng
no dual gap
MILP

Step 1: let $x^{k+1} \in \arg\min_{x \in X} L^1(x, z^k, u^k, \beta^k)$;

Step 2: $z^{k+1} \in \arg\min_{z \in Z} g^\top z$

$$+ \max_{j \in [k]} \{P(z^k, u^k, \beta^k) + \langle u^k, Bz - Bz^k \rangle - \beta^k \|Bz - Bz^k\|_1\};$$

Step 3: $u^{k+1} = u^k + \beta^k(Ax^{k+1} + Bz^{k+1})$;

$\beta^{k+1} = 1.1\beta^k$ when $\mathrm{mod}(k, 5) = 0$.

$$\parallel$$

Compare to original ADMM

Step 2: $z^{k+1} \in \arg\min_{z \in Z} g^\top z$

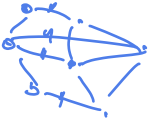
$$+ P(z^k, u^k, \beta^k) + \langle u^k, Bz - Bz^k \rangle + \beta^k \|Ax^{k+1} + Bz\|_1 - \|Ax^{k+1} + Bz^k\|_1$$

dual gap

Max Cut

Given $G = (V, E)$, the Max-Cut problem can be formulated as a MILP:

$$
\begin{aligned}
\min_{x,y} \quad & \sum_{(i,j)\in E} -y_{ij} \\
\text{s.t.} \quad & y_{ij} \leq x_i + x_j \\
& y_{ij} \leq 2 - x_i - x_j \\
& x_i, x_j, y_{ij} \in \{0,1\} \quad \forall (i,j) \in E
\end{aligned}
$$



$x_i \in \{0,1\}$

- $x_i = x_j \quad \Rightarrow \quad y_{ij} = 0 \quad \Rightarrow \quad (i,j)$ is not a cut
- $x_i \neq x_j \quad \Rightarrow \quad y_{ij} \leq 1 \quad \Rightarrow \quad$ objective minimized with $y_{ij} = 1$

Max Cut: Networks from *Matpower* with 3 partitions

β enough

			ALM			ADMM		
$\|V\|$	$\|E\|$	Couple	Iter.	Gap(%)	Time	Iter.	Gap(%)	Time
2383	2886	66	104(5)	1.14	59.11	31	1.46	17.27
2736	3496	80	58(4)	1.77	43.77	31	1.65	24.89
2746	3505	90	128(5)	2.07	82.80	31	1.67	21.33
2868	3471	54	71(5)	1.56	49.02	32	1.04	21.23
2869	3968	39	7(5)	0.42	11.18	12	0.48	11.76
3012	3566	72	91(5)	2.00	69.17	32	1.38	24.99
3120	3684	102	63(4)	2.11	40.70	31	1.71	20.86
3374	4068	77	35(4)	1.27	31.52	31	1.35	25.44
6468	8065	74	30(5)	1.09	54.09	13	0.80	21.89
6515	8104	71	15(5)	1.12	29.48	14	0.57	22.80
Avg		75.60	60.2(4.7)	1.45	47.08	**25.80**	**1.21**	**21.24**

Table: Comparison of ALM and ADMM

Randomly Generated MIPs

1. Parallel blocks: $X = \prod_{i=1}^{10} X_i$
   - $X_i = \{x \in [0,2]^{20} \times \{0,1\}^{30} \mid A_i x = b_i\}$
   - $A_i \in \mathbb{R}^{20\times 50}$ with standard Gaussian entries
   - $b_i = A_i x_0$, where x_0 is uniformly generated from $[0,2]^{20} \times \{0,1\}^{30}$
2. Equate some binary variables from *each pair* of blocks
3. Introduce z to break all couplings between $x_i's$ in step 2
4. Append z to an artificial master block

Randomly Generated MIPs

	Gurobi		ALM			ADMM		
case	Gap(%)	Time	Iter.	Gap(%)	Time	Iter.	Gap(%)	Time
1	NA	1200	143(27)	0.00	355.13	129	2.05	311.05
2	0.00	715.46	70(13)	0.00	109.67	62	2.65	**79.96**
3	NA	1200	114(25)	3.71	293.20	190	2.42	380.80
4	NA	1200	300(23)	NA	929.92	217	2.46	633.03
5	0.00	466.31	107(20)	4.24	**227.18**	218	2.08	527.66
6	NA	1200	74(22)	1.86	256.03	96	2.04	220.84
7	NA	1200	164(19)	2.26	930.30	180	4.96	954.10
8	NA	1200	150(28)	4.47	399.99	157	4.07	331.72
9	0.00	197.17	83(17)	2.78	**129.15**	113	2.58	159.94
10	NA	1200	57(15)	1.68	166.22	62	3.09	109.34
Avg			**126.2**(20.9)	**2.33**	379.68	142.4	2.84	**370.84**

Table: Comparison of ALM and ADMM

Dual Decomp

$\langle u, Ax + Bz \rangle$

Main references: $\underline{\underline{x}}$ $\underline{\underline{z}}$ $+ \| \ \|_{\rho}$

- 2-stage MILP: Benders'62
- ALM: Hestenes'69, Powell'69, analyzed by Rockafellar'73 and Bertsekas (see his book'14)
- ADMM: Gabay-Mercier'76, Glowinski-Marroco'75

  Glowinski LeTalloc '83

- Bound MILP exact penalty: Feizollahi-Ahmed-Sun'17

  $\rho < \frac{\sqrt{5}+1}{2}$

- Nonconvex ADMM: Wang-Y.-Zeng'19
- AL cut: Zhang-Sun'19, Ahmed-Cabral-da Costa'20
- Two-stage stochastic MILP: Benders'62, Van Slyke-Wets'69

  L-Shaped

- **MILP ADMM:** Sun-Sun-Y. arXiv'2102.11980