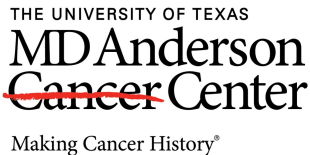


An Introduction to Alternating Direction Method of Multipliers (ADMM)

Yu Ding

The University of Texas MD Anderson Cancer Center

October 14th, 2024



Foundations and Trends® in
Machine Learning
Vol. 3, No. 1 (2010) 1–122
© 2011 S. Boyd, N. Parikh, E. Chu, B. Peleato
and J. Eckstein
DOI: 10.1561/22000000016



Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers

Stephen Boyd¹, Neal Parikh², Eric Chu³
Borja Peleato⁴ and Jonathan Eckstein⁵

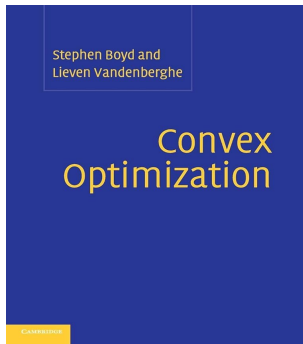
¹ *Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA, boyd@stanford.edu*

² *Computer Science Department, Stanford University, Stanford, CA 94305, USA, nparikh@cs.stanford.edu*

³ *Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA, echu508@stanford.edu*

⁴ *Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA, peleato@stanford.edu*

⁵ *Management Science and Information Systems Department and RUTCOR, Rutgers University, Piscataway, NJ 08854, USA, jeckstei@rci.rutgers.edu*



[1]

[2]

- 1 Lagrangian Duality
- 2 From Dual Asent to ADMM
- 3 Why ADMM
- 4 ADMM in CliPP

- 1 Lagrangian Duality
- 2 From Dual Asent to ADMM
- 3 Why ADMM
- 4 ADMM in CliPP

Primary Problem

$$\min_x f(x)$$

$$\text{s.t. } h_i(x) \leq 0, i \in \{1, \dots, I\}$$

$$\ell_j(x) = 0, j \in \{1, \dots, J\}$$

- The objective function $f(x)$ usually has poor properties (non-convex or no Lipschitz continuity).
- The feasible region is not always a convex set.

Lagrangian of the Primary Problem

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^I \lambda_i h_i(x) + \sum_{j=1}^J \nu_j \ell_j(x),$$

where λ, ν are Lagrangian multipliers.

Dual function

$$g(\lambda, \nu) = \inf_{x \in \text{dom} f} (f(x) + \sum_{i=1}^I \lambda_i h_i(x) + \sum_{j=1}^J \nu_j \ell_j(x)).$$

Note that, $g(\lambda, \nu)$ is concave w.r.t λ, ν , and the above infimum is taken over the domain of f .

A brief introduction to Lagrangian Duality

Primary Problem

$$\min_x f(x)$$

$$\text{s.t. } h_i(x) \leq 0, i \in \{1, \dots, I\}$$

$$\ell_j(x) = 0, j \in \{1, \dots, J\}$$

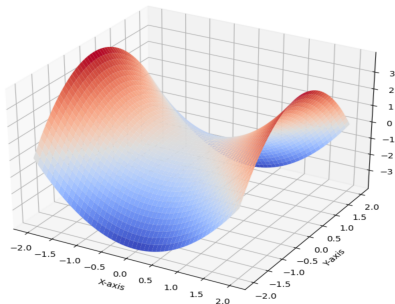
The dual problem is much easier to solve, as it is a convex optimization with linear constraints.

Dual Problem

$$\max_{\lambda, \nu} g(\lambda, \nu)$$

$$\text{s.t. } \lambda \geq 0.$$

Saddle Point Visualization: $f(x, y) = x^2 - y^2$



- 1 Lagrangian Duality
- 2 From Dual Asent to ADMM**
- 3 Why ADMM
- 4 ADMM in CliPP

Dual Ascent

Consider an easier problem $\min_x f(x)$ s.t. $Ax = b$.

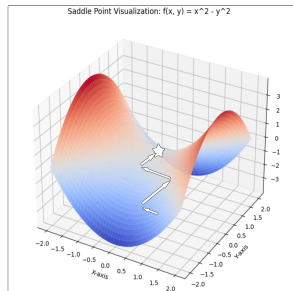
Its dual function is

$$g(y) = \inf_{x \in \text{dom}f} f(x) + y^T (Ax - b).$$

The gradient of $g(y)$ is $\frac{\partial g}{\partial y} = Ax^* - b$

The dual ascent algorithm is

- Initialize dual guess $y^{(0)}$
- repeat for $k = 1, 2, 3, \dots$
 - $x^{(k)} = \arg \min_{x \in \text{dom}f} f(x) + (y^{(k-1)})^T Ax$
 - $y^{(k)} = y^{(k-1)} + t_k (Ax^{(k)} - b)$



Augmented Lagrangian method a.k.a method of multipliers

The disadvantage of dual ascent is that it requires strong conditions to ensure convergence, i.e., convexity and Lipschitz continuity, etc. (It can be considered similar to the gradient descent algorithm.)

We propose a new primal problem

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|_2^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

The benefit that comes with the quadratic penalty term is that we can always adjust ρ such that the objective function is convex (under mild assumptions), as long as matrix A has full column rank.

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|_2^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- Initialize dual guess $y^{(0)}$
- repeat for $k = 1, 2, 3, \dots$
 - $x^{(k)} = \arg \min_{x \in \text{dom} f} f(x) + (y^{(k-1)})^T Ax + \frac{\rho}{2} \|Ax - b\|_2^2$
 - $y^{(k)} = y^{(k-1)} + t_k (Ax^{(k)} - b)$

Alternating Direction Method of Multipliers (ADMM)

Consider the problem

$$\begin{aligned} \min_{x,z} \quad & f(x) + g(z) \\ \text{s.t.} \quad & Ax + Bz = c \end{aligned}$$

The augmented Lagrangian is

$$L_\rho(x, z, u) = f(x) + g(z) + u^T (Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2$$

The algorithm is

- repeat for $k = 1, 2, 3, \dots$
 - $x^{(k)} = \arg \min_x L_\rho(x, z^{(k-1)}, u^{(k-1)})$
 - $z^{(k)} = \arg \min_z L_\rho(x^{(k)}, z, u^{(k-1)})$
 - $u^{(k)} = u^{(k-1)} + t_k (Ax^{(k)} + Bz^{(k)} - c)$

- 1 Lagrangian Duality
- 2 From Dual Asent to ADMM
- 3 Why ADMM**
- 4 ADMM in CliPP

Divide and Conquer

It can split a large problem into a series of subproblems. Usually, each subproblem has a closed form solution. Consider problem

$$\min_x f(x) + g(Ax)$$

We can transform it into

$$\min_x f(x) + g(z), \text{ s.t. } Ax - z = 0.$$

$$L_\rho(x, z, u) = f(x) + g(z) + u^T(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|_2^2$$

• repeat for $k = 1, 2, 3, \dots$

- $x^{(k)} = \arg \min_x g(z) + (u^{(k-1)})^T Ax + \frac{\rho}{2}\|Ax + Bz^{(k-1)} - c\|_2^2$
- $z^{(k)} = \arg \min_z f(x) + (u^{(k-1)})^T Bz + \frac{\rho}{2}\|Ax^{(k)} + Bz - c\|_2^2$
- $u^{(k)} = u^{(k-1)} + t_k(Ax^{(k)} + Bz^{(k)} - c)$

Distributed Optimization

Given $y \in \mathbb{R}^n$, $x \in \mathbb{R}^{n \times p}$, we have the group lasso problem as

$$\min_{\beta} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \sum_{g=1}^G c_g \|\beta_g\|_2$$

Rewrite as

$$\min_{\alpha, \beta} \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \sum_{g=1}^G c_g \|\alpha_g\|_2, \text{ s.t. } \beta - \alpha = 0.$$

ADMM steps:

- repeat for $k = 1, 2, 3, \dots$
 - $\beta^{(k)} = (X^T X + \rho I)^{-1} (X^T y + \rho(\alpha^{(k-1)} - \omega^{(k-1)}))$
 - for $g = 1, \dots, G$ do in parallel
 - $\alpha_g^{(k)} = R_{c_g \lambda / \rho}(\beta^{(k)} + \omega_g^{(k-1)})$
 - $\omega^{(k)} = \omega^{(k-1)} + \beta^{(k)} - \alpha^{(k)}$

- 1 Lagrangian Duality
- 2 From Dual Asent to ADMM
- 3 Why ADMM
- 4 ADMM in CliPP**

The problem, eq.(8), in CliPP is

$$\min_{\omega} -\ell(\omega) + \sum_{1 \leq i < j \leq s} \rho_{\lambda}(|\omega_i - \omega_j|).$$

Define $\eta_{ij} = \omega_i - \omega_j$, we can rewrite it as

$$\min_{\omega, \eta} -\ell(\omega) + \sum_{1 \leq i < j \leq s} \rho_{\lambda}(|\eta_{ij}|).$$

Then the augmented Lagrangian is

$$\begin{aligned} L(\omega, \eta, \tau, \lambda) = & -\ell(\omega) + \sum_{1 \leq i < j \leq s} \rho_{\lambda}(|\eta_{ij}|) + \frac{\alpha}{2} \sum_{1 \leq i < j \leq s} (\eta_{ij} - \omega_i + \omega_j)^2 \\ & - \sum_{1 \leq i < j \leq s} \tau_{ij}(\eta_{ij} - \omega_i + \omega_j). \end{aligned}$$

- repeat for $k = 1, 2, 3, \dots$
 - eq.(S3) $\boldsymbol{\omega}^{(k)} = (\mathbf{B}^T \mathbf{B} + \alpha \boldsymbol{\Delta}^T \boldsymbol{\Delta})^{-1} [\alpha \boldsymbol{\Delta}^T (\boldsymbol{\eta}^{(k-1)} - \boldsymbol{\tau}^{(k-1)}) - \mathbf{B}^T \mathbf{A}]$
 - eq.(S4) $\eta_{ij}^{(k)} = \arg \min_{\eta_{ij}} \frac{\alpha}{2} (\delta_{ij} - \eta_{ij})^2 + p_\lambda (|\eta_{ij}|)$
 - eq.(S5) $\boldsymbol{\tau}^{(k)} = \boldsymbol{\tau}^{(k-1)} - \alpha (\boldsymbol{\Delta} \boldsymbol{\omega}^{(k)} - \boldsymbol{\eta}^{(k)})$

<https://github.com/wyylab/CliPP/blob/master/src/kernel.cpp>

- Update ω
- Update η
- Update τ

```
254     linear = DELTA * (alpha * eta_old + tau_new) - B.cwiseProduct(A);
255
256     Minv = 1.0 / ((B.cwiseProduct(B)).array() + double(no_mutation) * alpha);
257     Minv_diag = Minv.asDiagonal();
258
259     trace_g = -alpha * Minv.sum();
260     if(!isnan(trace_g)){
261         std::cout << "Lambda: " << Lambda << "\titeration: " << k << "\ttime" << std::endl;
262         return -1;
263     }
264
265
266     Minv_outer = Minv * (Minv.transpose());
267
268     inverted = Minv_diag.array() - 1.0 / (1.0 + trace_g) * (-alpha) * Minv_outer.array();
269     w_new = inverted * linear;
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

Thank you!
Questions?



Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al.

Distributed optimization and statistical learning via the alternating direction method of multipliers.

Foundations and Trends® in Machine learning, 3(1):1–122, 2011.



Stephen Boyd and Lieven Vandenberghe.

Convex optimization.

Cambridge university press, 2004.