Augmented Lagrangian function with backtracking

Bang Cong Vu[1], Dimitri Papadimitrou[1]
3nLab & BeRC, Leuven, Belgium
bangcvvn@gmail.com, dpapadimitriou@3nlab.org

Mots-clés : *Backtracking, Lagrangian function, nonlinear constrained*

1 Introduction

This paper develops a numerical method based on the backtracking line search for solving the following nonlinear constrained problem. Let $h\colon \mathbb{R}^d \to \mathbb{R}$ and $c\colon \mathbb{R}^d \to \mathbb{R}^m$ $(m \leq d)$ be smooth functions ($\mathcal{C}^1$ at least), with μ_h and μ_c-Lipschitz gradient. Let C be a closed convex subset of $\mathbb{R}^d$, the problem is to

$$\text{minimize } h(x) \quad \text{subject to } c(x) = b, x \in C, b \in \mathbb{R}^m. \tag{1}$$

The problem (1) models various nonlinear optimization problems such as the max-cut [6], generalized eigenvalue [10], clustering [8], and factorization problem in SDP [2, 3, 5]. The augmented Lagrangian method (ALM) for the problem (1) implies

$$\begin{cases} x_{k+1} \in \operatorname{argmin}_{u \in C} \mathcal{L}(x, y_k, \rho_k) := h(x) + \langle x \mid y_k \rangle + \frac{\rho_k}{2}\|c(x) - b\|^2 \\ y_{k+1} = y_k + \sigma_k(c(x_{k+1}) - b). \end{cases} \tag{2}$$

Finding x_{k+1} from (2) is the crucial step. Even when c is linear, the closed-form expression of x_{k+1} is not available. In the seminal work [2], authors have chosen a first-order limited memory BFGS approach that employs a strong Wolfe-Powell line search, and the number of limited memory BFGS updates that are stored equals three. Recently, attempts to relax (2) and several inexact Augmented Lagrangian methods for constrained nonconvex problems have been investigated in [7, 13]. The former [7] applies a general version of ALM with an aggressive updating rule for the penalty parameter to nonconvex optimization problems with both equality and inequality constraints. The latter [13] applies a proximal framework to nonconvex problems with nonlinear equality constraints. Beside these efforts, the framework proposed in [9, 1, 11] can be applied to solve Problem (1). However, the primal step of [9] and [1] requires a subsolver and is not explicit, respectively. In addition, the algorithmic framework in [1] uses some additional conditions as in [1, Definition 3.1] and the boundedness of $(y_k)_{k\in\mathbb{N}}$. Lastly, the algorithm in [11] uses a inner loop to determine the penalty parameters $(\rho_k)_{k\in\mathbb{N}}$.

The objective of this paper is to develop an algorithmic framework based on the backtracking line search for the (augmented) Lagrangrian function associated to the problem (1). We propose an efficient numerical method relying on this algorithmic framework to solve problem (1).

2 Algorithm and Convergence

We propose the following algorithm where the dual update is the same as in (2) but the primal steps are obtained by using the backtracking line search on the Lagrangian function $\varphi_k = \mathcal{L}(\cdot, y_k, \rho_k)$ under the following conditions : for $Z = C$ and $Y = c(Z) - b$,

$$\mu_0 = \sup_{x\in C} \|\nabla c(x)^*\| < +\infty \text{ and } (\exists \zeta \in \,]0, +\infty[)(\forall (x, v) \in Z \times Y)\ \|\nabla c(x)^* v\| \geq \zeta \|v\|. \tag{4}$$

The convergence of the proposed method is characterized in terms of the gradient mapping, feasibility, and objective function as well as the convergence to stationary points.

Theorem 1 *Suppose that $(\mathcal{L}(x_k, y_k, \rho_k))_{k\in\mathbb{N}}$ is bounded below, $\|\nabla c(x_{k+1})^\star v_{k+1} - \nabla c(x_k)^\star v_k\| \leq \mu_c t_{k-1}\|y_{k-1}\|\|d_{k-1}\|$, and (4) is satisfied. Then $(x_{k+1} - x_k)/\sqrt{t_k} \to 0$ and $c(x_k) - b \to 0$. Moreover, under the additional qualification condition at $\overline{x}$: $-\nabla c(\overline{x})^* \overline{y} \in N_C(\overline{x}) \Longrightarrow \overline{y} = 0$, and $\inf_{k\in\mathbb{N}} t_k > 0$, every cluster point $\overline{x}$ of $(x_k)_{k\in\mathbb{N}}$ is a stationary point to the problem (1), i.e. there exists $\overline{y} \in \mathbb{R}^m$ such that $c(\overline{x}) = b$, and $-\nabla h(\overline{x}) - \nabla c(\overline{x})^* \overline{y} \in N_C(\overline{x})$. If $(\rho_k, x_k)_{k\in\mathbb{N}}$ is bounded and all assumptions hold then $(h(x_k))_{k\in\mathbb{N}}$ is a convergent sequence.*

3 Numerical Results

Among others, we execute by means of Octave code the Algorithm 1 for the SDP relaxation of the dual of the primal max-cut problem : $\min \langle C \mid P \rangle$ subject to $\operatorname{diag}(P) = 1,\ P \succeq 0$ on data sets [4]. This SDP relaxation is a particular case of (1) with $h \mapsto \langle C \mid P \rangle$, $c\colon P \mapsto \operatorname{diag}(P)$ and $b = 1$. The stop

Algorithm 1 ALM algorithm with backtracking

Initialization $x_0 \in C$, $y_0 \in \mathbb{R}^m$, $x_{-1} \neq x_0$ $\sigma_{-1} \gg 1$, $\rho_{-1} > 0$, $\epsilon \in]0, \mu_h/2[$.
for $k \leftarrow 0, n$ **do**
  Select $\rho_k \in]0, \infty[$ such that

$$\xi_k := \frac{\mu_h}{2} - \frac{\rho_k}{2}\|\nabla c(x_k)\|^2 > \varepsilon > 0, \rho_k\|c(x_k) - b\| \leq \sigma_{k-1}\|x_k - x_{k-1}\|, \ \rho_k - \rho_{k-1} < \epsilon\sigma_{k-1}$$

  Compute $v_k = y_k + \rho_k(c(x_k) - b)$ and $d_k = -(\nabla h(x_k) + \nabla c(x_k)^* v_k)\mu_h^{-1}$.
  Find $t_k > 0$ and $\sigma_k \geq \sigma_{k-1}$ such that

$$\begin{cases} \varphi_k(P_C(x_k + t_k d_k)) < \varphi_k(x_k) + t_k \Delta\varphi_k(x_k, d_k) + \mathcal{O}((k+1)^{-1-\epsilon}) \\ 8(1+\epsilon)t_k\Big(\zeta^{-1}\big(\mu_0\sigma_k + \mu_c\|y_k\|\big)\Big)^2 \leq \sigma_k\xi_k \end{cases} \tag{5}$$

  Update $x_{k+1} = P_C(x_k + t_k d_k)$ and $y_{k+1} = y_k + \sigma_k(c(x_{k+1}) - b)$
end for

condition is declared when the relative gap against the value found by cvx-Matlab solver is $< 10^{-5}$. Computational times, number of iterations and optimal values are averaged over 20 executions. The results are presented in the following table.

Dataset	**Optimal values**			**#Iterations**			**Computation Time**		
	Our	**NLS**	**FOAL**	**Our**	**NLS**	**FOAL**	**Our**	**NLS**	**FOAL**
G1	-12083.172	-12083.174	-12083.185	2498	9124	7970	5.637	20.959	22.487
GD97b	-15340.105	-15340.096	-15340.017	5656	4082	4805	0.944	1.068	1.018
LFAT5t	-189.881	-189.883	-189.882	306170	689870	901120	69.405	129	159.22
Tr200b	-1006.603	-1006.610	-1006.606	870	2125	1826	0.216	0.768	0.731
Tr20b	-48.667	-48.667	-48.667	753	498	315	0.124	0.310	0.267
Tr500	-3014.463	-3014.463	-3014.463	1749	2314	3123	0.642	1.134	1.378
Sbbrail	-834.049	-834.046	-834.056	2882	3543	5599	0.837	1.354	1.885
Sherman1	-279.532	-279.536	-279.533	4584	6407	7235	1.569	2.627	2.589

These results are then compared against those obtained with FOAL [11] and NLS-SDP [12] algorithm. As it can be observed, for every data set, our Algorithm 1 performs much better than both FOAL and NLS-SDP in terms of computational time. The same conclusion can be drawn for the number of iterations except for the data sets GD97b and Tr20b.

Références

[1] J. Bolte, S. Sabach, M. Teboulle, Nonconvex Lagrangian-based optimization : monitoring schemes and global convergence, *Mathematics of Operations Research*, Vol.43, pp.1210-1232, 2018.

[2] S. Burer and R. Monteiro, A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization, *Math. Program.*, Vol. 95, pp. 329-357, 2003.

[3] S. Burer and R. Monteiro, Local minima and convergence in low-rank semidefinite programming, *Math. Program.*, Vol. 103, pp. 427-444, 2005.

[4] https ://networkrepository.com/networks.php & https ://data.sbb.ch/pages/home20/

[5] M. De Santis, F. Rendl and A. Wiegele, Using a factored dual in augmented Lagrangian methods for semidefinite programming, *Operations Research Letters*, Vol. 46, pp. 523-528, 2018.

[6] M. X. Goemans, D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *Journal of the ACM*, Vol. 42, pp. 1115-1145, 1995.

[7] G. N. Grapiglia and Y.-X. Yuan, On the complexity of an augmented lagrangian method for nonconvex optimization, *IMA Journal of Numerical Analysis*, Vol. 41, pp. 1546-1568, 2021.

[8] D. G. Mixon, S. Villar and R. Ward, Clustering subgaussian mixtures by semidefinite programming, *https ://arxiv.org/pdf/1602.06612.pdf.*

[9] M. F. Sahin, et al., An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints, *Proc. of NeurIPS 2019*, pp. 13943-13955, 2019.

[10] J. Song, P. Babu and D. P. Palomar, Sparse generalized eigenvalue problem via smooth optimization, *IEEE Transactions on Signal Processing*, Vol. 63, pp. 1627-1642, 2015.

[11] B. C. Vu, A. Alacaoglu, M. F. Sahin, A. Yurtsever and V. Cevher, A first-order augmented Lagrangian framework for nonconvex optimization with nonlinear constraints, *ICML 2018 Workshop*, 2018.

[12] B.C. Vu and D. Papadimitrou, A nonlinear splitting method for constrained semidefinite programming, Under revision, Oct. 2021.

[13] Y. Xie and S. J. Wright, Complexity of proximal augmented lagrangian for nonconvex optimization with nonlinear equality constraints, *Journal of Scientific Computing*, Vol. 86, pp. 1-30, 2021.