Sobre a Inexatidão no Método de Direções Alternadas[1]

Gabriel Passos* Prof^a^. Dr^a^. Sandra Augusta Santos

Resumo

Este trabalho apresenta tópicos selecionados de análise convexa e regularização e exibe um resultado sobre a convergência do método ADMM *(Alternating Direction Method of Multipliers)*. O método ADMM assemelha-se ao método dos multiplicadores, entretanto, por considerar as variáveis de maneira particionada e atualizá-las de forma encadeada, como no método de *Gauss-Seidel*, explora convenientemente a estrutura de vários problemas de otimização. Tal característica o destaca em relação a outros métodos, especialmente na solução de problemas de grande porte. O ponto forte do principal resultado de convergência do ADMM apresentado neste trabalho reside no fato de que os subproblemas do método podem ser solucionados de forma inexata, o que permite a abordagem de uma classe ampliada de problemas.

Palavras-chave: Convexidade; Convergência; ADMM.

1 Introdução

Solucionar problemas de otimização convexa é uma tarefa comum em áreas que envolvem aprendizado de máquina e processamento de imagem. Em geral, tais problemas envolvem grandes conjuntos de dados, o que nos força a desenvolver algoritmos inteligentes que aproveitem toda a estrutura do problema subjacente e ainda, que utilizem ao máximo a arquitetura dos computadores atuais. Um dos métodos amplamente utilizados para abordar problemas convexos especiais é o ADMM.

Em [3], os autores relacionam o ADMM com o método de *Douglas-Rachford* e desenvolvem uma versão generalizada para um problema particular. São desenvolvidas e testadas com dados reais em [4] duas versões do ADMM que permitem inexatidão na resolução dos subproblemas envolvidos. Em [2], é feito um tratamento geral sobre o ADMM, que cobre desde as definições básicas até problemas específicos e possíveis implementações. Destaca-se em [5] o resultado sobre a complexidade de iteração do ADMM, obtido analisando o método como uma instância do HPE *(Hybrid Proximal Extragradient)*. Uma abordagem inexata parcial do método ADMM é proposta em [7].

Este trabalho está organizado da seguinte maneira. Na Seção 2 são apresentados conceitos básicos sobre convexidade. Na Seção 3, o ADMM clássico é definido e a convergência do ADMM generalizado é estabelecida. Na Seção 4, o ADMM é aplicado em dois problemas adequados ao método.

2 Convexidade

Apresentamos nessa seção um breve estudo sobre a convexidade de conjuntos e funções. Um tratamento completo sobre teoria de convexidade pode ser encontrado em [1].

Definição 1 (Conjunto Convexo)**.** Um conjunto $\mathcal{C} \subset \mathbb{R}^n$ é dito convexo se

$$\alpha y + (1-\alpha)x \in \mathcal{C}, \quad \forall x, y \in \mathcal{C}, \quad \forall\ \alpha \in [0,1].$$

Para manter a consistência dos resultados, é conveniente adotar a convenção de que o conjunto vazio é um conjunto convexo.

[1]Este trabalho foi financiado pela FAPESP, processo 2019/15992-5.

Resumimos abaixo uma pequena parcela da teoria que envolve as funções convexas reais estendidas (funções que podem assumir os valores $+\infty$ e $-\infty$).

Definição 2 (Epígrafo)**.** Sejam $\mathcal{C} \subset \mathbb{R}^n$ um conjunto convexo não vazio e $f : \mathcal{C} \to [-\infty, +\infty]$. Definimos o epígrafo de f como o conjunto

$$\text{epi}(f) := \{(x, \xi) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \mathcal{C},\ \xi \in \mathbb{R},\ f(x) \leq \xi\}.$$

A seguir, apresentamos uma definição de função convexa aplicável às funções reais estendidas.

Definição 3 (Função Convexa)**.** Sejam $\mathcal{C} \subset \mathbb{R}^n$ um conjunto não vazio e $f : \mathcal{C} \to [-\infty, +\infty]$. Dizemos que f é uma função convexa se $\text{epi}(f)$ é um conjunto convexo.

Os casos degenerados de funções convexas reais estendidas para as quais f é idêntica a $+\infty$ ou $-\infty$ em todo seu o domínio são excluídos por meio da seguinte definição.

Definição 4 (Função Própria)**.** Sejam $\mathcal{C} \subset \mathbb{R}^n$ um conjunto não vazio e $f : \mathcal{C} \to [-\infty, +\infty]$. Dizemos que f é própria se $f(x) > -\infty$ para todo $x \in \mathcal{C}$ e ainda existe algum $x \in \mathcal{C}$ tal que $f(x) < +\infty$.

Agora, definiremos no contexto de análise convexa, o que vem a ser uma função fechada.

Definição 5 (Função Fechada)**.** Sejam $\mathcal{C} \subset \mathbb{R}^n$ um conjunto não vazio e $f : \mathcal{C} \to [-\infty, +\infty]$. Dizemos que f é uma função fechada se $\text{epi}(f)$ é um conjunto fechado.

Os conceitos de subgradiente e subdiferencial são apresentados a seguir.

Definição 6 (Subgradiente)**.** Seja $f : \mathbb{R}^n \to (-\infty, +\infty]$ uma função própria. Dizemos que $u \in \mathbb{R}^n$ é um subgradiente de f em x se

$$f(y) \geq f(x) + \langle u, y - x\rangle, \quad \forall\, y \in \mathbb{R}^n. \tag{1}$$

Define-se o subdiferencial de f em um ponto como o conjunto de todos os subgradientes de f em x e o denotamos por $\partial f(x)$.

3 O método ADMM

Considere o seguinte problema de otimização

$$\begin{aligned} \min_{x,w} \quad & f(x) + g(w) \\ s.a \quad & Ax + Bw = c, \end{aligned} \tag{P}$$

em que $f : \mathbb{R}^m \to (-\infty, +\infty]$ e $g : \mathbb{R}^n \to (-\infty, +\infty]$ são funções convexas, próprias e fechadas, $A \in \mathbb{R}^{l\times m}$, $B \in \mathbb{R}^{l\times n}$ e $c \in \mathbb{R}^l$. A função Lagrangiana aumentada do problema (P) com parâmetro $\lambda > 0$ é

$$\mathcal{L}_\lambda(x, w, p) := f(x) + g(w) + \langle p, Ax + Bw - c\rangle + \frac{\lambda}{2}||Ax + Bw - c||^2.$$

Um dos métodos apropriados para tratar o problema (P) é o ADMM *(Alternating Direction Method of Multipliers)*, que tenta combinar a separabilidade do problema (P) com a convergência do método dos multiplicadores [2]. A proposta do ADMM é gerar, a partir de pontos iniciais x^0, w^0 e p^0, sequências (x^k), (w^k) e (p^k) que seguem um processo iterativo dado por

$$\begin{aligned} x^{k+1} &\in \arg\min_{x\in\mathbb{R}^m} \mathcal{L}_\lambda(x, w^k, p^k), \\ w^{k+1} &\in \arg\min_{w\in\mathbb{R}^n} \mathcal{L}_\lambda(x^{k+1}, w, p^k), \\ p^{k+1} &:= p^k + \lambda(Ax^{k+1} + Bw^{k+1} - c). \end{aligned} \tag{2}$$

Uma das dificuldades em utilizar o processo dado em (2) é que os subproblemas em x e w devem ser resolvidos de forma exata. A proposta do principal resultado abordado neste trabalho estabelece a convergência de uma versão mais geral do método ADMM que permite que os subproblemas sejam solucionados de forma inexata.

Baseados nas condições de *Kuhn-Tucker* para o problema (P) e no desenvolvimento do método ADMM feito em [3], definimos o conceito de tripla *Kuhn-Tucker*.

Definição 7 (Tripla *Kuhn-Tucker*)**.** Dizemos que $(\overline{x}, \overline{w}, \overline{p}) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l$ é uma tripla *Kuhn-Tucker* para o problema (P) se $(\overline{x}, \overline{w})$ satisfaz a igualdade $A\overline{x} + B\overline{w} = c$ e ainda

$$0 \in \partial f(\overline{x}) + A^t\overline{p}, \quad 0 \in \partial g(\overline{w}) + B^t\overline{p}.$$

Mostra-se que se $(\overline{x}, \overline{w}, \overline{p})$ é uma tripla *Kuhn-Tucker* para o problema (P), então $(\overline{x}, \overline{w})$ é solução de (P) e $\overline{p}$ solução do problema dual de (P).

Finalmente, apresentamos o resultado que garante a convergência do método ADMM generalizado, que segue a apresentação de [3]. As demonstrações dos próximos resultados podem ser encontradas em [6].

Teorema 1. *Considere o problema* (P) *e sejam dados* $x^0 \in \mathbb{R}^m$, $w^0 \in \mathbb{R}^n$, $p^0 \in \mathbb{R}^l$ *e* $\lambda \in \mathbb{R}$ *com* $\lambda > 0$. *Considere também as sequências* (μ_k), (ν_k) *e* (ρ_k) *de números reais tais que:*

- $\mu_k, \nu_k \geq 0$ *para todo* k *e* $\sum \mu_k, \sum \nu_k < +\infty$.
- $\rho_k \in (0, 2)$ *para todo* k *e* $0 < \liminf \rho_k \leq \limsup \rho_k < 2$.

Tome sequências (x^k), (w^k) *e* (p^k) *tais que para todo* k:

- $||x^{k+1} - \tilde{x}^k|| \leq \mu_k$.
- $||w^{k+1} - \tilde{w}^k|| \leq \nu_k$.
- $p^{k+1} = p^k + \lambda(\rho_k A x^{k+1} - (1-\rho_k)(Bw^k - c) + Bw^{k+1} - c)$,

em que

$$\tilde{x}^k \in \underset{x \in \mathbb{R}^m}{\arg\min} \left\{ f(x) + \langle p^k, Ax \rangle + \frac{1}{2}\lambda||Ax + Bw^k - c||^2 \right\},$$

$$\tilde{w}^k \in \underset{w \in \mathbb{R}^n}{\arg\min} \left\{ g(w) + \langle p^k, Bw \rangle + \frac{1}{2}\lambda||\rho_k A x^{k+1} - (1-\rho_k)(Bw^k - c) + Bw - c||^2 \right\}.$$

Então, se (P) *possui uma tripla Kuhn-Tucker, valem as afirmações*

1) p^k *converge para uma solução do problema dual de* (P).

2) $Ax^k + Bw^k - c \to 0$.

3) $f(\tilde{x}^k) + g(\tilde{w}^k) \to \overline{\tau}$.

4) Toda subsequência convergente (x^{k_j}, w^{k_j}) *tem como ponto limite uma solução do problema primal* (P).

Demonstração. Veja [6, Teorema 3.3.1] ∎

Corolário 1. *Se* f *e* g *são funções reais, então,* $f(x^k) + g(w^k) \to \overline{\tau}$.

Demonstração. Veja [6, Corolário 3.3.1] ∎

Corolário 2. *Suponha que as matrizes A e B possuam posto coluna completo. Dessa forma, se o problemas* (P) *admite uma tripla Kuhn-Tucker, então, as sequências (x^k) e (w^k) convergem para uma solução do problema primal* (P). *Caso contrário, ao menos uma das sequências (p^k) e (v^k) é ilimitada, em que $v^k := c - Bw^k$ para todo k.*

Demonstração. Veja [6, Corolário 3.3.2] ■

4 Problemas Estruturados

Apresentaremos a seguir dois problemas que possuem estruturas adequadas para a aplicação do método ADMM. Iniciaremos com uma abordagem ADMM para problemas convexos genéricos, dados por

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & f(x) \\ s.a \quad & x \in \mathcal{C}, \end{aligned}$$

em que $f : \mathbb{R}^m \to (-\infty, +\infty]$ é uma função convexa, própria e fechada e $\mathcal{C} \subset \mathbb{R}^m$ é um conjunto convexo, fechado e não vazio. Para transformar esse problema na forma do problema (P), levamos as restrições ao objetivo via função indicadora $\Phi_{\mathcal{C}}$ do conjunto $\mathcal{C}$, definida por

$$\Phi_{\mathcal{C}}(x) := \begin{cases} 0, & \text{se } x \in \mathcal{C} \\ +\infty, & \text{se } x \notin \mathcal{C}. \end{cases}$$

Nesse caso, temos

$$\begin{aligned} \min_{\substack{x\in\mathbb{R}^m \\ w\in\mathbb{R}^m}} \quad & f(x) + \Phi_{\mathcal{C}}(w) \\ s.a \quad & x - w = 0. \end{aligned}$$

No contexto do Teorema 1, tem-se

$$\begin{aligned} \tilde{x}^k &\in \arg\min_{x\in\mathbb{R}^m} \left\{ f(x) + \langle p^k, x\rangle + \frac{\lambda}{2}||x - p^k||^2 \right\}, \\ \tilde{w}^k &:= \mathrm{Proj}_{\mathcal{C}}(\rho_k x^{k+1} + (1-\rho_k)w^k + \frac{1}{\lambda}p^k), \\ p^{k+1} &:= p^k + \lambda(\rho_k x^{k+1} + (1-\rho_k)w^k - w^{k+1}), \end{aligned}$$

em que $\mathrm{Proj}_{\mathcal{C}}(y)$ denota a projeção de y no conjunto $\mathcal{C}$.

O problema Lasso *(Least Absolute Shrinkage and Selection Operator)* é dado por

$$\min_{x\in\mathbb{R}^m} \quad \frac{1}{2}||Mx - b||^2 + \alpha||x||_1,$$

em que α é um escalar positivo. Na forma do problema (P), o problema Lasso pode ser descrito como

$$\begin{aligned} \min_{x\in\mathbb{R}^m} \quad & \frac{1}{2}||Mx - b||^2 + \alpha||w||_1 \\ s.a \quad & x - w = 0. \end{aligned}$$

No contexto do Teorema 1, temos o processo iterativo dado por

$$\begin{aligned} x^{k+1} &:= (M^tM + \lambda I)^{-1}(M^tb + \lambda w^k - p^k), \\ w_i^{k+1} &:= \mathcal{S}_{\frac{\alpha}{\lambda}|y|}(\rho_k x_i^{k+1} + (1-\rho_k)w^k + \frac{1}{k}p^k), \quad \forall\, i = 1, ..., m, \\ p^{k+1} &:= p^k + \lambda(\rho_k x^{k+1} + (1-\rho_k)w^k - w^{k+1}), \end{aligned}$$

em que $\mathcal{S}_\gamma(y)$ é operador limiar suave (*soft thresholding*), isto é

$$\mathcal{S}_\gamma(y) := \begin{cases} y - \gamma, & \text{se } y \geq \gamma, \\ 0, & \text{se } |y| < \gamma, \\ y + \gamma, & \text{se } y \leq -\gamma. \end{cases}$$

5 Considerações Finais

Estudamos as propriedades fundamentais dos operadores monótonos e monótonos maximais, com o intuito de compreender melhor a convergência do método ADMM. Esse estudo resultou na demonstração da convergência de uma versão generalizada do método. Além disso, alguns problemas estruturados foram analisados no contexto da generalidade incluída ao método. O roteiro completo do estudo e as demonstrações que fizemos integram a monografia [6]. Esse projeto foi muito enriquecedor, já que tópicos como convexidade, regularização e teoria de operadores monótonos não fazem parte de disciplinas do curso de Matemática Aplicada e Computacional.

Referências

[1] D. P. Bertsekas. *Convex Optimization Theory.* Belmont: Athena Scientific, 2009.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato e J. Eckstein. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". Em: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122.

[3] J. Eckstein e D. P. Bertsekas. "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". Em: *Mathematical Programming* 55.1-3 (1992), pp. 293–318.

[4] J. Eckstein e W. Yao. "Approximate ADMM algorithms derived from Lagrangian splitting". Em: *Computational Optimization and Applications* 68.2 (2017), pp. 363–405.

[5] R. D. C. Monteiro e B. F. Svaiter. "Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers". Em: *SIAM Journal on Optimization* 23.1 (2013), pp. 475–507.

[6] G. Passos. "Convergência do algoritmo ADMM e aplicações a problemas estruturados". Monografia disponível em `https://www.ime.unicamp.br/~mac/db/2020-1S-172351.pdf`. Jun. de 2020.

[7] B. F. Svaiter. "A partially inexact ADMM with o(1/n) asymptotic convergence rate, #(1/n) complexity, and immediate relative error tolerance". Em: *Optimization* (2020), pp. 1–20.