Identificación de Clumps en Nubes Moleculares: Un Enfoque Desde el Cálculo Variacional

Martín Villanueva A.[1]

Descripción de Propuesta

Uno de los problemas fundamentales existentes hoy en astronomía, es la correcta identificación de clumps, y el análisis estructural de nubes moleculares, almacenadas en cubos espectroscópicos de datos.

Hace no mucho tiempo esta tarea era realizada manualmente por astrónomos, pero esto sólo era posible, pues trabajaban con imágenes pequeñas (con un par de canales en el espectro). Sin embargo en la actualidad, gracias a proyectos como ALMA, las imágenes generadas son de alta resolución, con un espectro en frecuencias más fino, haciendo prácticamente imposible su procesamiento manual. La motivación de los métodos computacionales, es poder automatizar estas tareas de manera correcta y eficiente.

Existen varias técnicas computacionales enfocadas a esta labor, dentro de las más notables se encuentran ClumpFind, GaussClumps[1] y FellWalker[2], todos ellas implementadas en una biblioteca llamada CUPID[3]. Sin embargo, estas implementaciones son dependientes de una gran cantidad de parámetros (no intuitivos de configurar), y al mismo tiempo son computacionalmente costosas.

Lo que se propone en el siguiente proyecto, es la elaboración de un modelo matemático haciendo uso del cálculo variacional, que permita desarrollar un algoritmo que supla las falencias anteriores, esto es; Que sea dependiente de pocos parámetros (e intuitivos de configurar por quien los utiliza), y computacionalmente eficiente. Adicionalmente se propone implementar tal algoritmo dentro del paquete de software ACALib[2] (*Advanced Computational Astronomy Library*)[4], para así finalmente realizar un estudio comparativo, con los algoritmos estándar en detección de clumps astronómicos.

Hipótesis

El cálculo variacional (o cálculo de variaciones) es una técnica que ha sido ampliamente utilizada tanto en matemáticas, ciencias (Física y Biología), como en aplicaciones ingenieriles. Algunos de los ejemplos más notables donde ha sido utilizada con éxito son: Modelado de estructuras en proteinas[5], Procesamiento de señales e imágenes [6], Machine learning [7], entre muchos otros [8]. En particular, en el ámbito del procesamiento digital de imágenes, permite realizar procesos como reconstrucción, restauración, segmentación, extracción de ruido y detección de patrones [6]. Como la identificación de clumps es una tarea que presenta evidentes similitudes con los problemas anteriores, este proyecto plantea como hipótesis, que un enfoque variacional es adecuado para el problema propuesto, permitiendo generar un análisis teórico/matemático mucho más preciso que el de los algoritmos de CUPID.

Siendo un poco más precisos, la hipótesis es que es posible definir un funcional Φ, tal que su mínimo sea una solución coherente del problema; Sea *f* la función que modela la distribución inherente en el cubo espectroscópico, y *g* la función que pretende aproximar tal distribución, un primer acercamiento al funcional en cuestión viene dado por:

$$\Phi(g) = \int_{\Omega \subset R^3} L(x, y, z, g, g_x, g_y, g_z) d\Omega$$
$$= \int_{\Omega \subset R^3} F_{\text{similitud}}(f, g) + \alpha F_{\text{penalización}}(f, g) + \beta\, F_{\text{suavidad}}(g_x, g_y, g_z)\; d\Omega$$

1 Departamento de Informática, UTFSM, Abril 2016.

2 https://github.com/ChileanVirtualObservatory/ACALIB Software desarrollado por LIRAE/UTFSM bajo el proyecto FONDEF IT15I10041

el cual posee tres términos de interés:*Similitud* que obtiene un mínimo cuando las funciones se asemejan más de acuerdo a cierta métrica, *Penalización* para generar soluciones consistentes a las restricciones, y *Suavidad* para filtrar el ruido presente en el cubo de datos.

Siguiendo los resultados del cálculo variacional, el problema de minimización se reduce a la resolución de la siguiente EDP:

$$\frac{\partial L}{\partial g} - \frac{d}{dx}\frac{\partial L}{\partial g_x} - \frac{d}{dy}\frac{\partial L}{\partial g_y} - \frac{d}{dz}\frac{\partial L}{\partial g_z} = 0 \quad \forall (x, y, z) \in \Omega$$

para la cual debe desarrollarse un método eficiente de resolución, y así cumplir con los objetivos del proyecto.

Objetivos

- Generar un modelo matemático bajo el enfoque del cálculo variacional que sea consistente con el problema a resolver, que no requiera de una gran cantidad de parámetros a configurar, y que los parámetros a usar posean una interpretación simple.
- Investigar los distintos métodos para la resolución del modelo anterior, así como las optimizaciones necesarias para generar un método computacionalmente eficiente, escalable y con posibilidades de paralelización.
- Implementación del método propuesto dentro de ACALib, por medio del lenguaje de programación Python, haciendo uso de bibliotecas optimizadas para métodos numéricos (NumPy, SciPy, Cython, Numba, entre otras).
- Análisis comparativo de los resultados obtenidos por el método aquí propuestos, con los algoritmos estándar implementados en CUPID[3]: FindClump, GaussClumps y FellWalker.

Metodología

La metodología utilizada para la realización del proyecto en sus distintas fases, se detalla a continuación:

- Reuniones semanales con profesor a cargo (Ph.D Claudio Torres), y con investigador a cargo (Ph.D Mauricio Araya) para discusión de problemas que se presenten, establecer directrices del proyecto, y obtener retroalimentación.
- Trabajo personal en temas del proyecto (modelamiento, programación, pruebas, documentación, entre otros). Por lo menos 4 bloques semanales de trabajo presencial en laboratorio CSRG F-119.
- Presentación de resultados al final de cada una de las cinco etapa del proyecto (detallados en **Plan de Trabajo**), con ambos profesores encargados.
- Modelo de desarrollo iterativo, con sistema de documentación de avances (bitácora), y uso de sistema de control de versiones GIT, de modo que profesores puedan mantenerse al tanto de los avances del proyecto.

3 Página principal de CUPID

Plan de Trabajo

Tareas	Inicio (semana)	Término (semana)
1. **Construcción del modelo teórico**		
1.1 Definición del funcional a minimizar	1	2
1.2 Estudio de métodos para la resolución de la ecuación de Euler-Lagrange	3	8
1.3 Estudio de métodos de optimización para la solución anterior	9	10
1.4 Verificación de validez del modelo obtenido	11	12
2. **Implementación de algoritmos**		
2.1 Determinación de herramientas (bibliotecas) adicionales a utilizar	13	13
2.2 Configuración del entorno de desarrollo	13	13
2.3 Definición de clases, métodos, estructuras y casos de uso	14	14
2.4 Desarrollo de los componentes del algoritmo	15	21
2.5 Documentación de código y proyecto	22	22
3. **Testeo y revisión de modelo**		
3.1 Verificación de interfaces y casos de uso	23	23
3.2 Incorporación de mejoras o modificaciones al modelo	24	26
4. **Análisis y comparación de resultados**		
4.1 Selección de datos y pruebas a realizar	27	27
4.2 Realización de pruebas, generación de gráficos y resultados comparativos	28	29
5. **Desarrollo de paper con resultados obtenidos**		
5.1 Estudio y realización del estado del arte	30	31
5.2 Desarrollo del fundamento teórico y del modelo generado	32	33
5.3 Desarrollo del funcionamiento del algoritmo	34	35
5.4 Presentación y análisis de resultados	36	38

Observación: Para la determinación de los periodos de tiempo se utilizan semanas, considerando las 39 semanas existentes entre el 11 de abril de 2016, al 11 de enero de 2017, equivalentes a los 10 meses de duración del proyecto.

Trabajo Adelantado

El equipo de LIRAE (Laboratory of Interdiciplinary Research on Astro-Engineering) ha venido desarrollando un paquete de software llamado ACALib (*Advanced Computational Astronomy Library*), cuyo objetivo es implementar un paquete de software, coherente con la investigación actual en los métodos computacionales en astronomía. Dentro de las funcionalidades implementadas hasta ahora en ACALib están:

- Rutinas de manipulación y tratamiento de cubos de datos espectroscópicos (Rotación, escalamiento y estandarización).
- Procedimientos para realizar stacking.
- Implementaciones propias y funcionales, de los algoritmos de identificación de clumps de CUPID.
- Técnicas de asociación de líneas espectrales.

- Un módulo para la generación de cubos de datos sintéticos.
- Varias rutinas para la visualización de gráficos 2D y 3D por medio de Matplotlib y Mayavi.

Contar con tales funcionalidades ya implementadas es una gran ventaja, pues proveen abstracciones, interfaces, y rutinas de manejo y visualización de datos, que permitirán centrar los esfuerzos del presente proyecto, en los temas que realmente le conciernen.

Recursos Disponible

Los recursos de los que se dispone actualmente para el desarrollo y puesta en marcha del proyecto, se listan a continuación:

- Espacio físico de trabajo en Laboratorio CSRG (Computer System Research Group), Departamento de Informática, F-119.
- Paquete de Software ACALib, que provee funcionalidades avanzadas para la manipulación de cubos espectroscópicos de datos y su correcta visualización (entre otras). Las funciones de manipulación, permiten presentar los datos de una manera estandarizada y fácil de usar. Las funciones de visualización ayudan a verificar los resultados de tales algoritmos.
- Workstation para HPC (High Performance Computing): Intel(R) Xeon(R) 8 core CPU X5647 @ 2.93GHz, con tarjeta gráfica Nvidia Tesla C2050 / C2070. (Perteneciente a grupo CSRG).

Referencias

[1] Stutzki and Güsten. 1990. CO and CS Observations on M17 SW. Appendix.

[2] Berry. 2014. FellWalker – a Clump Identification Algorithm.

[3] Berry. 2013. CUPID – A 3D Clump identification and Analysis Package. Starlink User Note 255.

[4] Araya et al. 2015. The ChiVO Library: advanced computational methods for astronomy.

[5] Papst. 2010. A Biological Application of the Calculus of Variations.

[6] Wang, Serpedin and Qaraqe. 2014. Variational Methods in Signal and Image Processing.

[7] Jordan, Ghahramani, Jaakkola and Saul. 1999. An introduction to Variational Methods for Graphics Models.

[8] Ferguson. A Brief Survey of the History of the Calculus of Variations and its Applications.