Propuesta de Tesis en Ingeniería Matemática

Identificación del proponente

- Nombre:
  Carlos López Vázquez (CLV)
  Raúl Tempone (RT)
- Último título obtenido:
  CLV: PhD, KTH, Estocolmo, Suecia
  RT: PhD, KTH, Estocolmo, Suecia
- Lugar de trabajo:
  CLV: Universidad Politécnica de Madrid/The Digital Map Ltda. – Montevideo
  RT: KAUST University, Arabia Saudita
- Área de trabajo:
  CLV: Calidad de Datos, Métodos Numéricos, etc. Ver http://www.thedigitalmap.com/~carlos
  RT: Métodos numéricos, ecuaciones diferenciales estocásticas.
- Información de contacto:
  carlos.lopez@thedigitalmap.com

Identificación de la propuesta de proyecto de tesis

Título del proyecto: Estimación de funciones R^2-->R^2 con restricciones y con un número insuficiente de datos

Área temática del conocimiento de la propuesta: Análisis numérico y matemático, Estadística y Geoestadística, Mecánica de los Fluidos

Resumen:
El problema de estimar un campo (escalar o vectorial) dados valores experimentales del mismo irregularmente distribuidos en el dominio es clásico en ciencias aplicadas, y en particular en Ciencias de la Tierra. Existen muchas fórmulas diseñadas para interpolar los valores funcionales; algunas de ellas no tienen parámetros libres (ej: interpolación ponderada con el inverso del cuadrado de la distancia) y otras tienen algunos pocos (krigeado). Pocas de estas fórmulas toman en consideración la existencia de restricciones adicionales (como ser producir campos de signo positivo, o con gradientes acotados, o preservar la correlación espacial, etc.). También fallan al ofrecer un subespacio de funciones suficientemente rico, lo que indirectamente genera el inconveniente de no asegurar que al aumentar el número de puntos dato haya convergencia a la función desconocida. El resultado neto es que el analista tiene pocas herramientas para lograr una función *adecuada.*

Un procedimiento estándar para medir lo adecuado de la función estimada es el de validación cruzada. El mismo consiste en subdividir los datos en dos conjuntos A y B, estimar la función utilizando solamente el conjunto A y aplicarla a las coordenadas del conjunto B. El mejor método será aquel en que minimiza cierta norma de las discrepancias. El carácter estadístico de la elección del "mejor" método ha sido analizado en López y González (2009), pero es un tópico recurrente en la literatura: una y otra vez los métodos son comparados entre sí generando un orden de precedencia entre ellos.

Dependiendo del problema hay otras restricciones duras que aplicar, por lo que la búsqueda puede afinarse incorporando restricciones al campo solución dada una estimación inicial arbitraria. En el caso de flujos, Tempone (1999) resuelve el problema incorporando la condición de divergencia nula; López (2010) logrando un campo derivado de un potencial, y seguramente hay otros. En estos casos el espacio de la solución está restringido indirectamente a través de la selección de una solución inicial. El conocido método de Krigeado (ver por ejemplo Myers, 2010) en el contexto de simulación resuelve el problema de incorporar información de correlación espacial sin restringir el espacio de búsqueda, pero nos deja sin expresión funcional para el

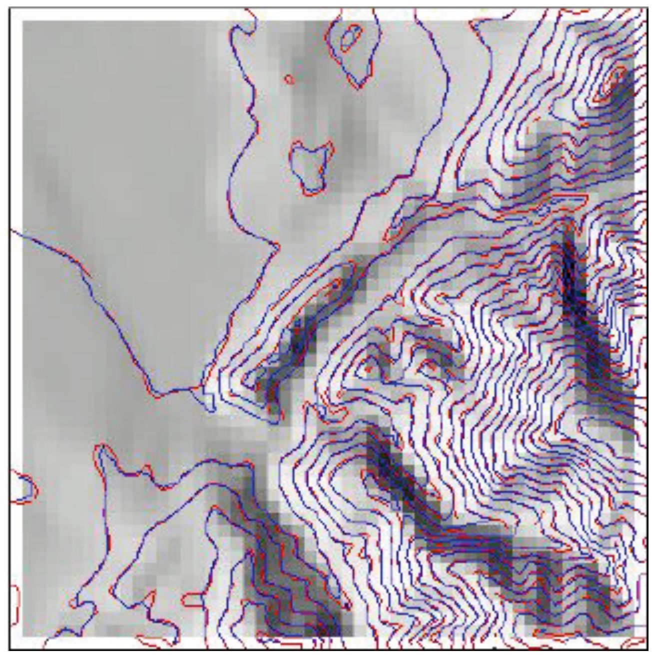
mismo. Las aplicaciones a registro de imágenes médicas (Noblet *et al.*, 2005), metamorfosis, conflación de cartografía (Casado, 2006), etc. requieren invertibilidad, preservación del sentido, topología, etc. En aplicaciones de geodesia se requiere además preservación de ángulos, por lo que se usan transformaciones conformes.

Es necesario conciliar la verificación exacta o aproximada de restricciones con la búsqueda de una función adecuada que replique la información disponible en un conjunto finito de puntos dato. El procedimiento será adecuado cuando se logre demostrar teóricamente (o inferir experimentalmente) la convergencia a la función solución al aumentar el número de datos. La métrica de éxito será siempre alguna norma de la discrepancia contra un conjunto de datos que no participan en la estimación de la función.

Posibles aplicaciones científicas:

El problema es estándar en ciencias experimentales, y el enfoque propuesto es en particular muy interesante para las ciencias duras, en las que hay restricciones fundamentales que pueden aplicarse como ser ecuaciones diferenciales. La búsqueda en Google Académico con "*interpolation scattered survey*" arroja casi 44.000 hits; la primer página de resultados ya es ilustrativa de la importancia asignada al tema, incluyendo referencias de publicaciones como *Mathematics of computation, Constructive Approximation, Journal of Computational and Applied Mathematics, ACM Transactions on Mathematical Software*, etc.

Posibles aplicaciones productivas y/o sociales: En Hidráulica y Mecánica de Fluidos podría mencionarse la estimación de campos de velocidades en laboratorio o de viento horario en el terreno con objetivos de prospección de energía eólica. En general, cualquier variable cuantitativa con coordenadas espaciales y variación continua requiere procedimientos como el señalado, tanto en R^2-->R^2 como R^2-->R (ej.: Modelo Digital de Elevación (ver figura), concentración de nutrientes en el terreno, etc.). La fusión o conflación de datos espaciales, de fuentes diferentes y que deben llevarse a un mismo marco geométrico requieren de transformaciones como las descritas. La operación de *registro* de imágenes médicas, localizando órganos homólogos en dos imágenes de distinta fecha permite analizar cuantitativamente los cambios observados (dilatación, contracción, etc.) y actuar en consecuencia.



Metodología:

Ya se ha realizado una recopilación sobre métodos clásicos de interpolación, así como una evaluación comparativa de su desempeño en una aplicación cartográfica. Los métodos corrientes típicamente ignoran restricciones del problema, preocupándose únicamente por replicar la información en los puntos dato, y dan una solución sin otros grados de libertad (i.e. perteneciente a un subespacio muy limitado). La línea de investigación que se propone es la de partir de un subespacio adecuado e imponer progresivamente las restricciones propias del problema en estudio, algunas estrictamente y otras aproximadamente. Del subespacio final se adoptará la solución más simple posible, aplicando criterios de Máxima Parsimonia u otros similares si es que hiciera falta.

Para su aplicación cartográfica, Casado (2006) establece que la función a interpolar es la que relaciona puntos homólogos entre dos cartografías. Así, dados los objetos homólogos (puntos, poligonales, polígonos, etc.) se intenta estimar una función de transformación $f(x,y)=(u,v)$ tal que $(X,Y)=(x+u,y+v)$ siendo (X,Y) las coordenadas transformadas del punto (x,y) y $u=u(x,y); v=v(x,y)$ funciones escalares. Típicamente estas funciones se eligen de un cierto subespacio, y los parámetros libres se estiman de forma de cumplir exacta o aproximada con algunas condiciones, como ser

$$(X_i, Y_i) = \left(x_i + u\left(x_i, y_i\right), y_i + v\left(x_i, y_i\right)\right), i = 0,1,\ldots,N:$$

o hacer mínima la suma S, definida como

$$S = \sum_{i=0}^{N} \left[(X_i, Y_i) - \left(x_i + u\left(x_i, y_i\right), y_i + v\left(x_i, y_i\right)\right)\right]^2$$

La primer expresión corresponde a un problema de Interpolación, mientras que la segunda define un problema de Aproximación. Si los datos (X, Y) no pueden asumirse sin error, entonces es más apropiado no obligar a la transformación a respetarlos exactamente sino aproximadamente, por lo que la segunda formulación parece más adecuada. Será la asumida.

En la mayoría de las aplicaciones son habituales las condiciones de continuidad para u y v. Si se impone además que la transformación es biunívoca aparecen restricciones duras al Jacobiano. Hay también restricciones blandas, que pueden ser del tipo "*...las áreas serán preservadas todo lo posible...*" que pueden ser incorporadas en términos de regularización de Tikhonov. Cada problema tendrá un contexto, y no puede esperarse encontrar una solución válida para todos ellos.

Un enfoque posible es el ilustrado en el trabajo de Tempone: generar una función interpolante y buscar una perturbación pequeña en cierto sentido que permita imponer sucesivamente restricciones duras (con multiplicadores de Lagrange) y blandas (como términos de regularización). La función interpolante debería pertenecer a un subespacio de partida suficientemente amplio, el cual (por ejemplo) podría modelarse con elementos finitos y funciones base de tipo C^1. La solución más general del problema (desde el punto de vista matemático + software) podría tener una biblioteca de restricciones "corrientes", y activarlas o no una a una según sea el caso. La solución numérica podría implicar llegar a una ecuación diferencial (caso de Tempone), resolver directamente una Optimización de gran escala en la formulación de partida, u otros enfoques que resultarán del trabajo. Al involucrar datos experimentales será inevitable manejar un enfoque estadístico de la solución, así como de la bondad del ajuste obtenido.

La actividad de tesis requerirá relevar la bibliografía en el tema, explorando áreas del conocimiento diversas (desde cartografía hasta imágenes médicas pasando por metamorfosis y animación digital), implementar (o recopilar) rutinas, realizar ensayos comparativos con todos los métodos ya existentes, quizá proponer alguna variante y ensayarla, etc. Al final debería elaborarse un documento de estado del arte y una biblioteca de rutinas accesible desde Octave/Matlab®.

*Bibliografía relevante (*indique entre 4 y 8 referencias)

López, C. y González, C.H. 2009 "Comparación de algoritmos para la conflación geométrica de información vectorial". VI Jornadas Técnicas de la Infraestructura de Datos Espaciales de España JIDEE'09, Murcia, España, ISBN 978-84-87138-56-0

Myers, D. 2010 http://www.ic.arizona.edu/ic/math574/ accedido 20101120

Noblet, V.; Heinrich, Ch.; Heitz, F. and Armspach, J.P. 2005 "3-D Deformable Image Registration: A Topology Preservation Scheme Based on Hierarchical Deformation Models and Interval Analysis Optimization" IEEE Transactions on Image Processing, 14, 5, 553

Tempone, R. 1999 "Approximation and interpolation of divergence free flows". Tesis de Maestría en Ingeniería Matemática, FING, UDELAR

Perfil esperado del estudiante: Conocimientos en Cálculo Numérico, con al menos un curso básico específico; familiaridad de programación en Matlab; conocimientos de métodos estadísticos básicos y disposición para considerar otros más avanzados; manejo de conceptos de Optimización y ecuaciones diferenciales parciales.

Lugar y Fecha de la propuesta: Montevideo, 19 de noviembre de 2010