



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

ICS3153 — Optimización Avanzada

Proyecto de Curso - Entrega II

Introducción

El problema de compressed sensing se trata de procesar y reconstruir una señal a partir de una muestra donde sólo se nos entrega parte de la información. Para esto, busca soluciones a sistemas lineales indeterminados. Un gran campo de aplicación de este problema es en el análisis y reconstrucción de imágenes, donde ha tenido un gran desarrollo. En este proyecto buscaremos resolver de forma eficiente una variación de este problema.

Tomemos un video estático cualquiera. Es decir que el fondo permanece relativamente fijo y casi sin movimiento por la mayor parte del tiempo. Supongamos que dentro del video existe un movimiento específico en algún lugar de la imagen (por ejemplo, un hombre cruza una calle). Si los cuerpos en movimiento dentro del video no ocupan una gran cantidad del espacio de la imagen, buscamos separar el fondo estático del primer plano, es decir, los cuerpos en movimiento del fondo fijo.



Figure 1: El primer cuadro de la figura muestra a un grupo de personas caminando en la calle, el segundo cuadro el fondo estático y finalmente los elementos en movimiento distiguídos del fondo.

Luego, para este problema, podemos usar la siguiente representación: buscamos matrices D , L y S tal que

$$D = L + S$$

donde D las columnas son los cuadros del vídeo, L representa el fondo y S el primer plano del vídeo [1], es decir el movimiento. Asumiendo una posición permanente del foco de la cámara, entonces se logra un fondo relativamente fijo, perturbado solo por condiciones naturales y cambios de luz.

Bajo estas condiciones, podemos asumir que la matriz S , correspondiente a las perturbaciones de la imagen, será suficientemente dispersa (sparse), ya que afecta solo a una pequeña proporción de las entradas

del fondo y la matriz L tendrá una alta dependencia entre sus columnas. De esta forma, esta última matriz deberá tener el menor rango posible (idealmente 1).

Luego nuestro problema de interés es recuperar la matriz L , dada una matriz $D = S + L$, con S y L desconocidas, pero considerando L del menor rango posible y S dispersa. Para encontrar las matrices de interés, se busca resolver el problema

$$\begin{aligned} \min_{L,S} \quad & \text{rank}(L) + \lambda \|S\|_0 \\ \text{s.a} \quad & D = L + S \end{aligned}$$

Sin embargo, no estamos enfrentando a un problema no convexo donde no existe una solución eficiente y por lo tanto un problema NP-Hard. Luego, es posible tratar una relajación del problema planteado reemplazando la norma 0 por norma 1 y el rango de la matriz con la norma nuclear definida por

$$\|A\|_* = \sum_i \sigma_i(A) = \text{trace}(\sqrt{A^T A})$$

Luego nuestro problema de interés se transforma a

$$\begin{aligned} \min_{L,S} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{s.a} \quad & D = L + S \end{aligned}$$

Observemos que $\|L\|_* + \lambda \|S\|_1$ es la envoltura convexa de $\text{rank}(L) + \lambda \|S\|_0$ sobre el conjunto de todas las matrices (L, S) tales que

$$\max(\|A\|_{2,2}, \|E\|_{1,\infty}) \leq 1$$

donde la norma $\|A\|_{\alpha,\beta}$ viene dada por

$$\|A\|_{\alpha,\beta} = \max_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}$$

De acuerdo con Wright et al. [2], es posible recuperar perfectamente la matriz de bajo rango L mediante la resolución del problema relajado.

Para resolver el problema que acabamos de presentar, tomaremos dos modelos distintos para aproximar la minimización de la cobertura convexa de nuestro problema. En la próxima sección, mostraremos un método eficiente para resolver el problema si es que no consideramos ningún tipo de relación entre los elementos distintos de cero de nuestra matriz de errores, es decir, no asumimos ninguna relación especial en nuestra matriz E además del hecho de ser sparse. Por su parte, luego mostraremos un modelo donde asumimos que los elementos distintos de cero son cercanos. Podemos asumir esto debido a que al tratarse de ruido de cuerpos moviéndose frente a un fondo, los píxeles que variarían en cada fotograma serían píxeles contiguos. Esta idea será profundizada en la tercera sección.

Finalmente, implementaremos ambos métodos y compararemos sus resultados. Esto, tanto en la velocidad de resolución de cada algoritmo como en el resultado obtenido al separar nuestros videos. Esto se realizará para la siguiente entrega.

Modelo matriz de errores sparse general

En primer lugar, consideraremos la matriz de errores como una matriz que solo será considerada sparse, sin ningún otro requisito para esta. Tal como vimos en la sección anterior, este problema puede ser resuelto de forma exacta (o suficientemente cercana) resolviendo simplemente la cobertura convexa de nuestro problema inicial. Es decir, nos basta resolver el problema

$$\begin{aligned} \min_{L, S} \quad & \|L\|_* + \lambda \|S\|_1 \\ \text{s.a.} \quad & D = L + S \end{aligned}$$

Para resolverlo, planteamos una resolución a través del método del langrangeano aumentado. Este método se utiliza para solucionar problemas del tipo $\min f(x) \quad \text{s.a.} \quad h(x) = 0$ con $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$. A los que uno define el langrangeano aumentado como

$$\mathcal{L}(X, Y, \mu) = f(x) + \langle Y, h(X) \rangle + \frac{\mu}{2} \|h(X)\|_F^2$$

con μ un escalar positivo. Luego, tomando una secuencia creciente $\{\mu_k\}$ obtenemos que el mínimo se obtiene con el siguiente algoritmo:

```

1: Sea  $Y^0, \mu_0$ 
2: while no ha convergido do
3:   Resolver  $X_{k+1} = \arg \min_X \mathcal{L}(X, Y^k, \mu_k)$ 
4:    $Y^{k+1} = Y^k + \mu_k h(X^{k+1})$ 
5:    $k = k + 1$ 
6: end while
7: Output:  $X^*$ 

```

Las ventajas que tiene utilizar este algoritmo es que ofrece una convergencia bastante rápida con condiciones poco exigentes. Tal como se demuestra en el paper de Lin et al. [3] nos basta con un problema continuo y sabemos que, teniendo nuestro problema, todo punto de acumulación (L^*, S^*) será un óptimo, y se cumplirá que $|\|L^*\|_* + \lambda \|S^*\|_1 - f^*| = O(\mu_{k-1}^{-1})$.

Por esto, sabemos que el algoritmo del langrangeano aumentado ofrece una muy buena convergencia a nuestro problema, además de ser de un bajo costo de resolución. Por todo esto es que es un método muy recomendable para la solución de este problema. Para ello tomamos $X = (L, S)$, $f(X) = \|L\|_* + \lambda \|S\|_1$ y $h(X) = D - L - S$ y su langrangeano es

$$\mathcal{L}(L, S, Y, \mu) = \|L\|_* + \lambda \|S\|_1 + \langle Y, D - L - S \rangle + \frac{\mu}{2} \|D - L - S\|_F^2$$

Sin embargo, todavía nos enfrentamos a un problema de grandes dimensiones, y obtener el resultado exacto de $\min \mathcal{L}(L, S, Y, \mu)$ sigue siendo demasiado costoso. Por ello, resolveremos una aproximación de este algoritmo, conocida como Langrangeano Aumentado Inexacto (LAI) donde obtendremos valores inexactos para cada paso de nuestra minimización. Esto implicará una mayor cantidad de iteraciones para obtener el resultado buscado, sin embargo cada iteración será menos costosa. Este algoritmo lo podemos escribir como:

```

1: Sea  $Y^0, S^0 = 0, \mu_0$ 
2: while no ha convergido do
3:   Resolver  $L_{k+1} = \arg \min_L \mathcal{L}(L, S^k, Y^k, \mu_k)$ 
4:   Resolver  $S_{k+1} = \arg \min_S \mathcal{L}(L^{k+1}, S, Y^k, \mu_k)$ 
5:    $Y^{k+1} = Y^k + \mu_k (D - L^{k+1} - S^{k+1})$ 
6:    $k = k + 1$ 
7: end while
8: Output:  $L^*, S^*$ 

```

Con el método LAI obtendremos respuestas de alta exactitud con tiempos de convergencia relativamente pequeños.

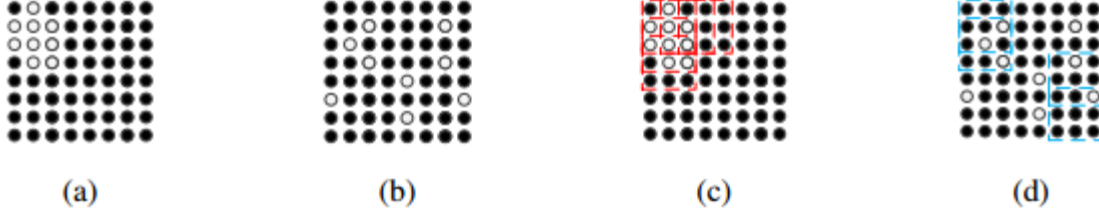


Figure 2: (a)-(b) Dos distribuciones sparse en un frame de 8x8. (c)-(d) Ejemplo de algunos cuadrados de 3x3 en cada caso con los que se calcularia la norma (1).

Modelo con errores sparse estructurados

Uno de los problemas presentes con la formulación anterior es que la norma l_1 no toma en cuenta la estructura o relación entre los componentes de la matriz en cuestión. Esto es un problema ya que en la práctica las cosas en el primer plano suelen ser objetos con cierta estructura más que solo ruido. Por lo mismo, en [1] se propone un método de que induce una matriz sparse ‘por grupos’. Este corresponde al método ‘Low-rank and Structured Sparsity Decomposition (LSD)’

Para este método es necesario definir la norma de ‘sparsity’ estructurada. Esta se define como

$$\Omega(S) = \sum_{j=1}^n \sum_{g \in \mathcal{G}} \|s_g^j\|_\infty \quad (1)$$

Donde $S \in \mathbb{R}^{m \times n}$ es la matriz del video es cuestión. La j -ésima columna de S , s^j , representa el frame j de forma extendida y \mathcal{G} contiene conjuntos de índices, donde cada uno es un subconjunto de los índices del frame. Entonces, s_g^j representa a los elementos del frame j que corresponden al conjunto de índices g . En el caso de [1], los índices se eligen de tal manera que estos representen cuadrados de 3x3 en el frame que se sobrelapan entre sí, de manera que cada cuadrado comparte 6 pixeles con su vecino. Entonces, lo que se hace es tomar en cuenta el pixel de mayor valor dentro de un cuadrado de 3x3.

La ventaja que ofrece la norma descrita en (1) es que esta favorece matrices en las cuales los elementos se encuentren más agrupados. De forma intuitiva, si una imagen tiene los valores más agrupados, entonces estos se considerarán en un menor número de cuadrados de 3x3. Un ejemplo de esto se puede ver en la Figura 2. En este caso las estructuras de 3a y 3b tendrán la misma norma l_1 , pero la norma (1) tendrá un valor menor en la distribución de 3a que en la de 3b. Esto se debe a que en la imagen existe un menor número de grupos que serán sumados. Se puede ver que los 5 cuadrados de 3x3 dibujados en 3c serán únicos con suma positiva, mientras que en la otra distribución el número de cuadrados con suma positiva será mucho mayor.

El modelo propuesto entonces se puede escribir como

$$\begin{aligned} \min_{L, S} \quad & \|L\|_* + \lambda \Omega(S) \\ \text{s.t.} \quad & D = L + S \end{aligned}$$

Para la resolución de este método es posible utilizar el lagrangeado del problema,

$$\mathcal{L}(L, S, Y, \mu) = \|L\|_* + \lambda \Omega(S) + \langle Y, D - L - S \rangle + \frac{\mu}{2} \|D - L - S\|_F^2$$

y utilizar los mismos métodos de Lagrangeano aumentado exacto e inexacto descritos en la sección anterior. Lo esperable es que los resultados obtenidos con la implementación de esta formulación se vean menos afectados por el ruido del vídeo y por variaciones pequeñas en el fondo y que, por lo mismo, los resultados obtenidos posean una mayor estructura.

References

- [1] Xin Liu, Guoying Zhao, Jiawen Yao, and Chun Qi. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Transactions on Image Processing*, 24(8):2502–2514, 2015.
- [2] Ganesh A. Rao S. Peng Y. Ma Y. Wright, J. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization. *In: Proceedings of Advances in Neural Information Processing Systems*, 2009.
- [3] Chen M. Ma Y. Lin, Z. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *In: NIPS*, 2011.