

結合  $\beta$  距離與圖形正規限制式之非負矩陣分解  
應用於單通道訊號源分離

Monaural Source Separation Using Nonnegative Matrix Factorization  
with Graph Regularization Constraint

林彥伯 Yan-Bo Lin

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[102502529@cc.ncu.edu.tw](mailto:102502529@cc.ncu.edu.tw)

范俊 Tuan Pham

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[103582605@cc.ncu.edu.tw](mailto:103582605@cc.ncu.edu.tw)

李遠山 Yuan-Shan Lee

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[102582003@cc.ncu.edu.tw](mailto:102582003@cc.ncu.edu.tw)

王家慶 Jia-Ching Wang

國立中央大學資訊工程學系

Department of Computer Science and Information Engineering

National Central University

[jcw@csie.ncu.edu.tw](mailto:jcw@csie.ncu.edu.tw)

摘要

本論文發展基於非負矩陣分解之單通道訊號源分離技術。有鑑於目前大多數非負矩陣分解方法，在計算成本函數(Cost Function)時多採用歐式距離(Euclidean Distance)或凱氏分歧度(Kullback-Leibler Divergence)等，而對於不同種類之未知訊號源，常因不同距離之選擇而造成分離效果有落差。因此，我們引入  $\beta$  距離進行單通道訊號源分離，藉由  $\beta$  之調控，使原本固定的距離選擇變為更加地彈性。同時，我們考量到，在利用非負矩陣分解進行訊號源分離時，混合訊號在高維度空間中隱含低維度平滑之流形(Manifold)分佈，因此我們將圖形正規限制式(Graph Regularization Constraint)導入最佳化問題中，藉

此在非負矩陣分解時，保留原來資料蘊含之幾何結構，來增強單通道訊號源分離的效果。

關鍵詞：非負矩陣分解，流形學習，訊號源分離，圖型正規

## 1. 緒論

非負矩陣分解(non-negative matrix factorization, NMF)[2]近年來被已廣泛應用於圖形處理[2,3]、音訊處理[10,12,13]等領域。在單通道訊號原分離方法方面，目前已有許多基於非負矩陣分解之方法出現[2,3,9,10,13,14]。大部分的方法在時頻域上進行分離，首先將單通道之混合訊號經過短時傅立葉轉換(STFT)至時頻域，接著以絕對值取出時頻訊號之能量(Magnitude)，形成混合訊號矩陣  $\mathbf{V} \in R_{F \times N}^+$ 。其主要目標是找到基底(Basis)矩陣  $\mathbf{W} \in R_{F \times K}^+$  與係數(Coefficients)矩陣  $\mathbf{H} \in R_{K \times N}^+$ ，使得混合訊號可以被表示為  $\mathbf{V} \approx \mathbf{WH}$ ，其最佳化成本函數(Cost Function)如下：

$$(\hat{\mathbf{W}}, \hat{\mathbf{H}}) = \arg \min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{X} \parallel \mathbf{WH}) \quad (1)$$

其中  $D(\parallel)$  為一任意測度(Metric)， $\mathbf{W}$ 、 $\mathbf{H}$  皆為非負矩陣， $\mathbf{W}$  的行向量為能量頻譜的基底成分， $\mathbf{H}$  的列向量則表示為對應成分在不同時間上的比例。在最佳化過程中通常使用迭代更新法，分別交互更新  $\mathbf{W}$  與  $\mathbf{H}$  之值，當(1)式中  $\mathbf{V}$  與  $\mathbf{WH}$  之測度如歐式距離(Euclidean Distance)、凱氏分歧度(Kullback–Leibler Divergence)，其值小於事先定義之門檻值時，則停止迭代。

由於不同的距離選用會得到不同的拆解結果，進而影響分離效果。為了改善這個問題，本論文引入了  $\beta$  距離( $\beta$ -divergence)於非負矩陣分解，進行單通道訊號源分離。藉由  $\beta$  之調控，相較於其他固定之距離，彈性地增加距離的可變動性。此外，有鑑於稀疏非負矩陣分解(Sparse Non-negative Matrix Factorization, SNMF)在單通道訊號源分離上已被證實有顯著的效果[5]-[8]，故我們額外加入稀疏限制式於最佳化成本函數中，強迫係數矩陣  $\mathbf{H}$  稀疏，以強化分離效果。

另一方面，為了在進行非負矩陣分解時，保留原來訊號之內蘊結構，我們假設混合訊號在高維度空間中隱含低維度平滑之流形(Manifold)分佈，加入了圖形正規限制式(Graph Regularization Constraint)於最佳化成本函數中，使得在原始混合訊號中彼此鄰近的兩點，在新的基底空間下亦相互鄰近，企圖進一步增強上述方法之分離效果。

## 二、運用 $\beta$ 距離非負矩陣分解與圖形正規限制式進行盲訊源分離

在本章節安排方面，我們在第一小節扼要回顧基於  $\beta$  距離之非負矩陣分解；在第二小節中，講述結合圖形正規限制式之非負矩陣分解；最後，在第三小節中說明基於  $\beta$  距離與圖形正規限制式之稀疏非負矩陣分解，及如何以其進行單通道訊號源分離。

### (一)、基於 $\beta$ 距離之非負矩陣分解

考量到在進行矩陣分解時，固定的量度(如歐氏距離、凱氏分歧度等)無法適應所有問題，

Févotte 等人[3]於 2011 年提出了基於  $\beta$  距離之非負矩陣及其演算法， $\beta$  為一個距離控制參數，藉由調整不同的  $\beta$  值，可使距離的選擇更加彈性。我們將其表示為  $d(x \parallel y)$ ，並定義  $\beta$  距離如下：

$$d_{\beta}(x \parallel y) \stackrel{def}{=} \begin{cases} \frac{1}{\beta(\beta-1)}(x^{\beta} + (\beta-1)y^{\beta} - \beta xy^{\beta-1}), & \beta \in \mathfrak{R} \setminus \{0,1\} \\ x \log \frac{x}{y} - x + y, & \beta = 1 \\ \frac{x}{y} - \log \frac{x}{y} - 1, & \beta = 0 \end{cases} \quad (2)$$

其中  $x \in \mathbf{R}_+$ ,  $y \in \mathbf{R}_+$  分別為一維的實數值，若  $\beta = 0$  則  $\beta$  距離退化為 IS 距離(Itakura-Saito Distance)，若  $\beta = 1$  則為凱氏分歧度，若  $\beta = 2$  則  $d_{\beta=2} = (x - y)^2$  恰為歐式距離。我們將(2)式對  $y$  微分，探討其特性：

$$\frac{\partial}{\partial y} d_{\beta}(x \parallel y) = y^{\beta-2}(y - x) \quad (3)$$

$$\frac{\partial^2}{\partial y^2} d_{\beta}(x \parallel y) = y^{\beta-3}[(\beta-1)y - (\beta-2)x] \quad (4)$$

我們會發現到會有一個最小值在  $y = x$  上，且其值會隨著  $|y - x|$  增加而遞增；同時也可以發現， $\beta \in [1,2]$  時此距離為為凹函數，其餘區間可以將式子分解如下：

$$d_{\beta}(x \parallel y) = \check{d}(x \parallel y) + \hat{d}(x \parallel y) + \bar{d}(x) \quad (5)$$

其中  $\check{d}(x \parallel y)$  為  $y$  的凹函數、 $\hat{d}(x \parallel y)$  為  $y$  的凸函數、 $\bar{d}(x)$  為常數。

當此分解  $\beta > 0$  時，可能包含歐式距離或凱氏分歧度，而不同的  $\beta$  值可能使此距離對於不同大小數值之敏感度不同。值得一提的是，IS 分歧度 ( $\beta = 0$ )，具有尺度不變(Scale Invariant)的特性，數學式子可表示為  $d_{is}(\lambda x \parallel \lambda y) = d_{is}(x \parallel y)$ ，其中  $\lambda$  為尺度純量。另外，當  $\beta$  為較小的正數時，對應之  $\beta$  距離較適用於的時頻訊號之處理。

在進行基於  $\beta$  距離之非負矩陣分解時，先假設  $\mathbf{W}$  固定，並定義  $C(\mathbf{H})$  為其成本函數，以最佳化方法更新  $\mathbf{H}$ ，其式子可表示如下：

$$\min_{\mathbf{H} \geq 0} C(\mathbf{H}) \stackrel{def}{=} D_{\beta}(\mathbf{V} \parallel \mathbf{W}\mathbf{H}) \quad (6)$$

其中  $\mathbf{V} \in \mathbf{R}_{F \times N}^+$ ， $\mathbf{W} \in \mathbf{R}_{F \times K}^+$ ， $\mathbf{H} \in \mathbf{R}_{K \times N}^+$ 。成本函數  $C(\mathbf{H})$  可分  $n$  個行向量之子問題  $\sum_n = D(\mathbf{v}_n \parallel \mathbf{W}\mathbf{h}_n)$ ， $\mathbf{v}_n$ 、 $\mathbf{h}_n$  分別為  $\mathbf{V}$ 、 $\mathbf{H}$  第  $n$  條行向量，我們將子問題簡寫表示如下：

$$\min_{\mathbf{h} \geq 0} C(\mathbf{h}) \stackrel{def}{=} D_{\beta}(\mathbf{v} \parallel \mathbf{W}\mathbf{h}) \quad (7)$$

其中  $\mathbf{v} \in \mathbf{R}_F^+$ 、 $\mathbf{W} \in \mathbf{R}_{F \times K}^+$ 、 $\mathbf{h} \in \mathbf{R}_K^+$ 。成本函數  $C(\mathbf{h})$  對  $\mathbf{h}$  微分後可形成梯度  $\nabla_{\mathbf{h}} C(\mathbf{h})$  如下式：

$$\nabla_{\mathbf{h}} C(\mathbf{h}) = \mathbf{W}^T [(\mathbf{W}\mathbf{h})^{(\beta-2)} (\mathbf{W}\mathbf{h} - \mathbf{v})] \quad (8)$$

最後，我們可以根據(8)式得出 $\mathbf{H}$ 與 $\mathbf{W}$ 之更新公式(Updating Rule)如下：

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^T [(\mathbf{W}\mathbf{H})^{(\beta-2)} \cdot \mathbf{V}]}{\mathbf{W}^T [\mathbf{W}\mathbf{H}]^{(\beta-1)}} \quad (9)$$

$$\mathbf{W} \leftarrow \mathbf{W} \cdot \frac{[(\mathbf{W}\mathbf{H})^{(\beta-2)} \cdot \mathbf{V}] \mathbf{H}^T}{[\mathbf{W}\mathbf{H}]^{(\beta-1)} \mathbf{H}^T} \quad (10)$$

其中“ $\cdot$ ”為矩陣元素點(Element-wise)運算。經過交互迭代更新基底 $\mathbf{W}$ 與係數 $\mathbf{H}$ 後，即完成基於 $\beta$ 距離之非負矩陣分解。

由於上述方法在矩陣分解時並無考慮資料隱含之幾何結構，故在下一小節中，我們將會討論以圖形正規限制式結合非負矩陣分解方法，在矩陣分解時保留原始資料的幾何結構。

## (二)、圖形正規限制式結合非負矩陣分解：

由於現實世界中的資料數據，通常以低維度流形(manifold)方式嵌在高維度的樣本空間中[1]，Cai等人[4]提出圖形正規化非負矩陣分解(Graph Regularized Non-negative Matrix Factorization, GNMF)將非負矩陣分解視作對資料的一種降維(Dimension Reduction)問題，結合圖形正規限制式使原資料投影後能夠保留其幾何性質。Kim等人[16]提出保留流形結構之音訊分離方法，在分離時透過保留時頻訊號之隱含幾何結構，亦即若原座標點 $\mathbf{x}_i$ 、 $\mathbf{x}_j$ 在原坐標系是鄰近的兩點，那麼在新的基底空間下亦為鄰近的兩點，來提升音訊分離的效果。

在本論文中，為了建立筆資料之關係，我們定義權重矩陣 $\mathbf{U}_{ij}$  (weight matrix)如下：

$$\mathbf{U}_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}} \quad (11)$$

其中 $t \in \mathbf{R}$ 為一距離參數，我們可以定義拉普拉斯矩陣(laplacian matrix) $\mathbf{L} = \mathbf{D} - \mathbf{U}$ ， $\mathbf{D}$ 為對角矩陣，其 $n$ 個元素計算方式為 $\mathbf{D}_{ii} = \sum_j \mathbf{U}_{ij}$ ，最後，我們將圖形限制 $R$ 定義如下：

$$\begin{aligned} R &= \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{x}_j - \mathbf{x}_i\|^2 \mathbf{U}_{ij} \\ &= \sum_{j=1}^N \mathbf{x}_j^T \mathbf{x}_j \mathbf{D}_{jj} - \sum_{j,i=1}^N \mathbf{x}_j^T \mathbf{x}_i \mathbf{U}_{ji} \\ &= \text{Tr}(\mathbf{H}\mathbf{D}\mathbf{H}^T) - \text{Tr}(\mathbf{H}\mathbf{U}\mathbf{H}^T) \\ &= \text{Tr}(\mathbf{H}\mathbf{L}\mathbf{H}^T) \end{aligned} \quad (12)$$

藉由最小化 $R$ ，我們可以觀察到再次觀察到前述的假設，若原座標點 $x_i$ 、 $x_j$ 在原本的坐標系是鄰近的兩點，那麼在新的基底空間下亦為鄰近的兩點。結合圖形正規限制式與原成本函數後，最佳化式子如下：

$$GNMF = \frac{1}{2} \| \mathbf{V} - \mathbf{WH} \|_F^2 + \frac{1}{2} \lambda \text{Tr}(\mathbf{HLH}^T) \quad (13)$$

在加入了圖形正規限制式後。非負矩陣分解可以考慮高維度資料在低維空間中的緊緻嵌入，並保留原始資料之幾何特性。

### (三)、結合 $\beta$ 距離與圖形正規限制式之稀疏非負矩陣分解：

本論文提出結合  $\beta$  距離與圖形正規限制式之稀疏非負矩陣分解，來解決單通道訊號源分離問題，除了使用  $\beta$  距離使距離選擇更加彈性以外，我們假設混合訊號在高維度空間中隱含低維度平滑之流形分佈，將圖形正規限制式導入最佳化問題中。此外，我們還額外加了針對係數矩陣  $\mathbf{H}$  之稀疏項，其最佳化式子如下：

$$(\hat{\mathbf{W}}, \hat{\mathbf{H}}) = \arg \min_{\mathbf{W}, \mathbf{H} \geq 0} D_\beta(\mathbf{V} \| \mathbf{WH}) + \mu \| \mathbf{H} \|_1 + \alpha \text{Tr}(\mathbf{HLH}^T) \quad (14)$$

其中  $\mu$  與  $\alpha$  分別控制  $l_1$  稀疏懲罰項與圖形正規限制項之權重。根據(9)、(10)式，我們可以推導出下列迭代更新式：

$$\mathbf{W} \leftarrow \mathbf{W} \cdot \frac{(\mathbf{A}^{\beta-2} \cdot \mathbf{V} + \tilde{\mathbf{W}} \tilde{\mathbf{W}} \mathbf{A}^{\beta-1}) \mathbf{H}^T}{(\mathbf{A}^{\beta-1} + \tilde{\mathbf{W}} \tilde{\mathbf{W}}^T (\mathbf{A}^{\beta-2} \cdot \mathbf{V})) \mathbf{H}^T} \quad (15)$$

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\tilde{\mathbf{W}}^T (\mathbf{V} \cdot \mathbf{A}^{\beta-2}) + \alpha \mathbf{H} \mathbf{U}}{\tilde{\mathbf{W}} \mathbf{A}^{\beta-1} + \alpha \mathbf{H} \mathbf{D} + \mu} \quad (16)$$

其中，“.”為矩陣元素點運算。矩陣  $\mathbf{A} = \tilde{\mathbf{W}} \mathbf{H}$ ，而  $\tilde{\mathbf{W}}$  則是  $\mathbf{W}$  經過行間正規劃 (Column-wise Normalization) 後得出。

接下來，我們介紹如何以所提出之非負矩陣分解方法，進行單通道訊號源分離，其分為訓練與分離兩階段。在訓練階段，我們的主要目的是找出音源之對應字典，假設有兩類型之音源，在做法上，我們蒐集用源訊號之時頻能量  $\mathbf{V}_{s1} \in R_+^{m \times n_1}$ 、 $\mathbf{V}_{s2} \in R_+^{m \times n_2}$ ，並利用(15)與(16)式迭代更新，求得對應的字典矩陣  $\mathbf{W}_{s1} \in R_+^{m \times k_1}$ 、 $\mathbf{W}_{s2} \in R_+^{m \times k_2}$ 。

在分離階段，我們將混合訊號經過短時傅立葉轉換轉換至時頻域得到  $\mathbf{V}_{mix} \in C^{m \times n}$  並取出其能量  $\bar{\mathbf{V}}_{mix} \in R_+^{m \times n}$ ，注意，如果訊號源類型是已知狀態下，我們將分離階段稱做監督式(Supervised)分離，監督型字典  $\mathbf{W}_{dic} \in R_+^{m \times (k_1 + k_2)}$  可以定義如下：

$$\mathbf{W}_{dic} = [\mathbf{W}_{s1} \quad \mathbf{W}_{s2}] \quad (17)$$

若是只知道部分音源類型，我們則稱其為半監督式(Semi-supervised)分離，我們可以假設(17)的部分字典為未知，並且利用(15)式從混合音源中學習出其他音源之字典。最後，我們可以透過以(17)式之字典，並透過(16)式得到混合音源對各源訊號字典之對應係數  $\mathbf{H}_{s1} \in R_+^{k_1 \times n}$  與  $\mathbf{H}_{s2} \in R_+^{k_2 \times n}$ ，並利用柔性遮罩(Soft Mask)完成分離：

$$\hat{\mathbf{V}}_{s1} = \bar{\mathbf{V}}_{mix} \cdot (\mathbf{W}_{s1} \mathbf{H}_{s1}) ./ (\mathbf{W}_{dic} \mathbf{H}_{mix}) \quad (18)$$

$$\hat{\mathbf{V}}_{s2} = \bar{\mathbf{V}}_{mix} \cdot (\mathbf{W}_{s2} \mathbf{H}_{s2}) ./ (\mathbf{W}_{dic} \mathbf{H}_{mix}) \quad (19)$$

其中  $\mathbf{H}_{mix} = [\mathbf{H}_{s1} \mathbf{H}_{s2}]^T$ ，“.”為矩陣元素點運算， $\hat{\mathbf{V}}_{s1}$ 與 $\hat{\mathbf{V}}_{s2}$ 為分離出之時頻訊號。我們的演算法可整理如下：

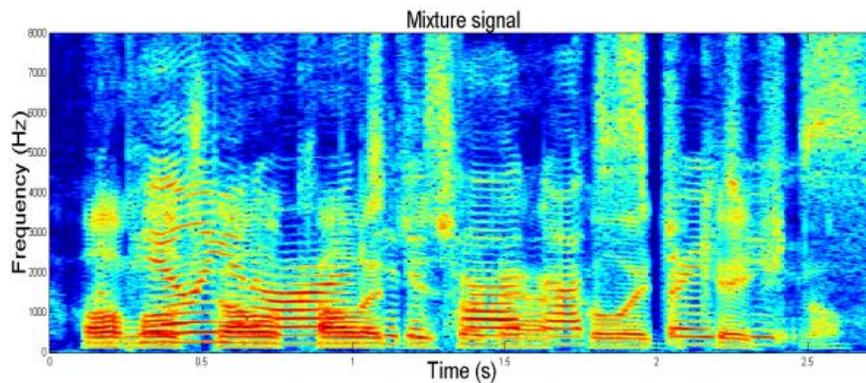
結合  $\beta$  距離與圖形正規限制式之稀疏非負矩陣分離演算法：

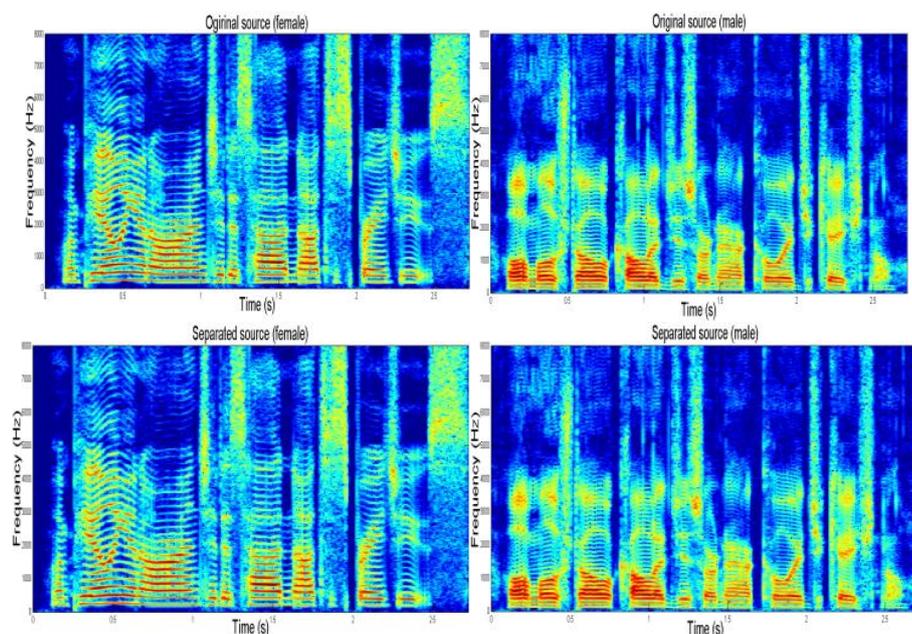
- 1：輸入： $\mathbf{V}_{mix} \in R_+^{m \times n}$ 、 $\mathbf{V}_{s1} \in R_+^{m \times n_1}$ 、 $\mathbf{V}_{s2} \in R_+^{m \times n_2}$  (訓練時  $V_{s1}, V_{s2}$  為輸入；否則  $\mathbf{V}_{mix}$ )
- 2：輸出： $\mathbf{W}_{dic}$  或  $\mathbf{H}_{mix}$  (訓練時，輸出為字典  $\mathbf{W}_{dic}$ ；在測試時，輸出係數矩陣  $\mathbf{H}_{mix}$ )
- 3：藉由隨機數值初始化缺失矩陣(Missing Matrix)
- 4：利用式子(11)建構與輸入符合的關連矩陣  $\mathbf{U}$ ，與對角矩陣  $\mathbf{D}$
- 5：**repeat**
- 6：    運用(16)更新  $\mathbf{H}_{mix}$
- 7：    **if** 訓練階段
- 8：        更新  $\mathbf{W}_{dic}$  運用(15)
- 9：    **end if**
- 10：**until** convergence

### 三、實驗結果

在實驗中，我們使用 TIMIT 資料庫[11]之男女聲進行分離實驗，取樣率為 16000 赫茲。我們隨機選取 4 名女性語者和 4 名男性語者，且每位語者念 10 個句子。我們對於每個語者隨機選擇一個句子建構混合訊號，其他 9 個句子作為訓練集。在評量方面，我們利用 BSS Eval toolbox[13]去計算 SDR(Signal-to-distortion Ratio)、SIR(Signal-to-interference Ratio)、SAR(Signal-to-artifact Ratio)三種評比。

本論文與稀疏非負矩陣分解[15]進行比較，我們採用與其相同的設定。共迭代 400 次，參數  $\alpha$  和  $\mu = 0.1$ 、 $\beta = 0$ 、元素(Atom)數  $K = 1024$ 。圖一為所提出之音源分離方法所得到之結果，我們在實驗中進行男聲與女聲之單通道音源分離。





圖一、單通道音源分離結果

在本實驗中，我們將所提出的方法與以下兩方法較：其一為稀疏非負矩陣分解 (NMF+S)、其二為基於  $\beta$  距離之稀疏非負矩陣分解(SNMF)[5]。

我們將分為兩種情況討論：

1)單一訓練：在這種狀況下，我們僅考慮一男一女語者去建構字典  $\mathbf{W} = [\mathbf{W}_1 \ \mathbf{W}_2]$ ，對於所有測試集均使用此字典，表一為五組測試句子的平均結果，實驗結果顯示，與傳統非負矩陣分解與稀疏非負矩陣相比，本論文所提出之方法具有較優之分離效果。在實際應用中，我們無法得知所有語者的字典，因此單一訓練之做法較為接近真實情況。

表一、單一訓練字典之單通道音源分離結果

	NMF+S	SNMF	本論文提出方法
SDR	3.50	7.11	<b>7.55</b>
SIR	5.39	11.97	<b>12.07</b>
SAR	9.37	9.19	<b>9.82</b>

2)全訓練：我們運用全部的語者去建構字典。由表二可以得知，三種演算法的效能都比顯然單一訓練時好。的確，當我們擁有較多的先驗資訊，則可得到較好的分離效果。在此實驗中，除了 SAR 略輸之外，其他評比皆優於稀疏非負矩陣分解方法。

表二、全訓練字典之單通道音源分離結果

	NMF+S	SNMF	本論文提出方法
SDR	5.14	8.93	<b>9.41</b>
SIR	7.74	13.46	<b>14.60</b>
SAR	9.92	<b>11.36</b>	<b>11.36</b>

觀察表一及表二之實驗結果可以得知，SNMF 相較於 NMF+S 有顯著的提升，而本論文所提出之方法，透過圖形正規限制式的加入，在三項評比皆優於 SNMF。

## 四、結論

本論文提出新穎之單通道訊號源分離方法，我們結合圖形正規限制式於與基於  $\beta$  距離之非負矩陣分解方法，藉由  $\beta$  之調控，使原本固定的距離選擇變為更加地彈性，而圖形正規限制式使得在非負矩陣分解時，保留原來資料蘊含之幾何結構，實驗結果顯示所提出的方法相較於稀疏非負矩陣分解，在分離效果方面有著顯著的提升。在未來展望方面，除了進行相關參數的最佳化(如  $\beta$  等)，我們還想進一步考慮資料具有群組的特性，在各群組內保留幾何結構，及考慮群組稀疏性。此外，由於訊號經由遮罩後容易產生失真的情形，我們也期望未來能透過後處理的方式，來強化經遮罩後的分離結果。

## 參考文獻

- [1] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, 15(6):1373-1396, 2003.
- [2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, Cambridge, MA, USA: MIT Press, vol. 13, 2001.
- [3] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the beta-divergence," *Neural Computation*, 2011.
- [4] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1548-1560, Aug. 2011.
- [5] J. L. Roux, F. Weninger, and J. R. Hershey, "Sparse NMF – half-baked or well done?," *Mitsubishi Electric Research Laboratories Technical Report*, Mar. 2015.
- [6] K. Minje and P. Smaragdis, "Mixtures of local dictionaries for unsupervised speech enhancement," *IEEE Signal Processing Letters*, vol. 22, no. 3, pp. 293 - 297, March 2015.
- [7] A. Lefèvre, F. Bach, and C. Févotte, "Itakura-Saito non-negative matrix factorization with group sparsity," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, 2011.
- [8] P. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457-1469, 2004.
- [9] B. King, C. Fevotte, and P. Smaragdis, "Optimal cost function and magnitude power for NMF-based speech separation and music interpolation," in *Proc. 2012 IEEE Int. Workshop on Machine Learning for Signal Processing*, 2012, pp. 1-6.
- [10] G. G. François and J. M. Gautham, "Nonnegative Matrix partial co-factorization for spectral and temporal drum source separation" *IEEE Signal Processing Letters*, vol. 21, no. 10, Oct. 2014.
- [11] S. Seneff, J. Glass V. Zue, "Speech database development at MIT: Timit and beyond," *Speech Communication*, vol. 9, no. 4, pp. 351-356, Aug. 1990.
- [12] G. Bao, Y. Xu, and Z. Ye, "Learning a discriminative dictionary for single-channel

- speech separation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 7, pp. 1130-1138, Apr. 2014.
- [13] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech and Language Processing, no. 14, pp. 1462-1469, 2006.
- [14] J. Eggert and E. Körner, "Sparse coding and NMF," IEEE International Joint Conference on Neural Networks, vol.4, pp. 2529-2533.
- [15] N. Mikkil, "Speech separation using non-negative features and sparse non-negative matrix factorization," Technical Reports, 2007.
- [16] M. Kim, P. Smaragdis and G. J. Mysore, "Efficient Manifold Preserving Audio Source Separation Using Locality Sensitive Hashing" IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP), 2015.