A Novel Dictionary Based Approach for Missing Sample Recovery of Signals in Manifold

Baburaj M. · **Sudhish N George**

Projection Operator

$$\mathscr{P}_{\mathscr{M}}(x) = \min_{\widetilde{x} \in \mathscr{M}} \|\widetilde{x} - x\| \tag{1}$$

Informally, given an arbitrary vector $x \in \mathbb{R}^{\mathbf{N}}$, the operator $\mathscr{P}_{\mathscr{M}}(x)$ returns the point on the manifold $\mathscr{M}$ that is closest to x, where closeness is measured in terms of the Euclidean norm. The projection operator $\mathscr{P}_{\mathscr{M}}(.)$ as euclidean projection onto $\mathscr{M}$. Note that in a number of applications, $\mathscr{P}_{\mathscr{M}}(.)$ as euclidean projection onto $\mathscr{M}$ may be quite difficult to compute exactly. Therefore, we define γ-approximate projection operator as

$$\widetilde{x} = \mathscr{P}^{\gamma}_{\mathscr{M}}(x) \implies \widetilde{x} \in \mathscr{M}, \text{ and } \|\widetilde{x} - x\| \leq \|\mathscr{P}_{\mathscr{M}}(x) - x\| + \gamma \tag{2}$$

so that $\mathscr{P}^{\gamma}_{\mathscr{M}}(.)$ yields a vector $\widetilde{x} \in \mathscr{M}$ that approximately minimizes the squared distance from x to $\mathscr{M}$.

Graident Operator

Let $\widetilde{x} \in \mathbb{R}^{\mathbf{N}}$ be the signal having arbitrary missing samples. The original clean signal $x \in \mathscr{M}$, so that, corrupted signal can be modelled as

$$\widetilde{x} = x + e, \;\; x \in \mathscr{M} \tag{3}$$

where e is the error due to missing samples. In worst case, noisy version of signal can be modelled as,

$$\widetilde{x}_{\eta} = x + e + \eta, \;\; x \in \mathscr{M} \tag{4}$$

where η is the additive white Gaussian noise present in the signal. If we project clean signal x onto $\mathscr{M}$, the result is signal x itself, whereas the projection of $\widetilde{x}$ yields $\hat{x} = \mathscr{P}_{\mathscr{M}}(\widetilde{x})$, $\check{x} \in \mathscr{M}$ & $\|\widetilde{x} - \check{x}\|_2 \leq \varepsilon$, where ε is a positive quantity. Let $p_i \in \mathbb{I}$ for $i = 1 \dots M$ are the locations of M missing samples. For each missing positions at k^{th} iteration we form two signals.
we form two signals.

$$\hat{x}^{(k)}_{+\delta}(n) = \begin{cases} \hat{x}^{(k)}(n) + \Delta & \text{if } n = p_i \\ \hat{x}^{(k)}(n) & \text{if } n \neq p_i. \end{cases}$$

$$\hat{x}^{(k)}_{-\delta}(n) = \begin{cases} \hat{x}^{(k)}(n) - \Delta & \text{if } n = p_i \\ \hat{x}^{(k)}(n) & \text{if } n \neq p_i. \end{cases}$$

where Δ is the step size. The projection of these signals onto manifold $\mathscr{M}$ is calculated as

$$\check{x}^{(k)}_{+\delta} = \mathscr{P}_{\mathscr{M}}(\hat{x}^{(k)}_{+\delta})$$

$$\check{x}^{(k)}_{-\delta} = \mathscr{P}_{\mathscr{M}}(\hat{x}^{(k)}_{-\delta})$$

The distance between the projected signal and original signal is measured as

$$d^{(k)}_1 = \|\hat{x}^{(k)}_{+\delta} - \check{x}^{(k)}_{+\delta}\|_2$$

$$d^{(k)}_2 = \|\hat{x}^{(k)}_{-\delta} - \check{x}^{(k)}_{-\delta}\|_2$$

The gradient is evaluated as

$$\mathscr{G}^{(k)}(p_i) = \frac{d^{(k)}_1 - d^{(k)}_2}{2\Delta} \qquad \forall i = 1 \dots M \tag{5}$$

$\mathscr{G}$ will have the same dimension as signal with missing samples $\widetilde{x} \in \mathbb{R}^{\mathbf{N}}$. The missing values of signal is corrected by applying correction.

Stability and Convergence

Analysis of stability and convergence of proposed method is presented. Stability have a direct dependence on the hyper-parameters and detailed discussion of hyper-parameters is presented in the *section 4*.

Theorem 1 *Let $x = (x_1, \dots x_j \dots x_n) \in \mathscr{M}$ with $x(j) \neq 0$ and $\widetilde{x} = (x_1, \dots \widetilde{x}_j \dots x_n) \in \mathscr{R}^n$ be a single sample corrupted version of x such that $\widetilde{x}_j = 0$ or $\widetilde{x}(j) = 0$. Then if $\hat{x}^{(k)}$ and $\hat{x}^{(l)}$ be two recovered versions of x such that $\|x - \hat{x}^{(k)}\| \leq \|x - \hat{x}^{(l)}\|$ then the respective gradients defined by* (5) *will be related as $\mathscr{G}^{(k)}(j) \leq \mathscr{G}^{(l)}(j)$.*

Proof From the definition or projection operator (1) we have $\|x - \mathscr{P}_{\mathscr{M}}(x)\| = 0$

$$\mathscr{G}^{(k)}(j) = \frac{\left\| \hat{x}^{(k)}_{+\delta} - \mathscr{P}_{\mathscr{M}}\left(\hat{x}^{(k)}_{+\delta}\right) \right\|_2 - \left\| \hat{x}^{(k)}_{-\delta} - \mathscr{P}_{\mathscr{M}}\left(\hat{x}^{(k)}_{-\delta}\right) \right\|_2}{2\Delta}$$

Without the loss of generality, we can assume that $x_j > 0$, $x^{(k)}_j \leq x_j$ and $x^{(j)}_j \leq x_j$ then

$$\mathscr{G}^{(k)}(j) \approx \frac{\sqrt{\left\{(\hat{x_j}^{(k)} + \Delta) - (x_j - \varepsilon_1)\right\}^2} - \sqrt{\left\{(\hat{x_j}^{(k)} - \Delta) - (x_j - \varepsilon_2)\right\}^2}}{2\Delta}$$

where $\varepsilon_1 > 0, \varepsilon_2 > 0$ & $\varepsilon_1 < \varepsilon_2$

$\mathscr{G}^{(k)}(j) \approx \frac{(\hat{x_j}^{(k)}+\Delta)-(x_j-\varepsilon_1)-(\hat{x_j}^{(k)}-\Delta)+(x_j-\varepsilon_2)}{2\Delta}$

$\therefore \mathscr{G}^{(k)}(j) \approx \frac{2\Delta+\varepsilon_1-\varepsilon_2}{2\Delta}$

$$\mathscr{G}^{(k)}(j) \approx 1 - \frac{\varepsilon_2 - \varepsilon_1}{2\Delta} \tag{6}$$

Select the parameter Δ such that $\frac{\varepsilon_2-\varepsilon_1}{2\Delta} \leq 1$.

Since $x_j^{(j)} \leq x_j$ and using the definition (6)

$\mathscr{G}^{(l)}(j) \approx 1 - \frac{\varepsilon_4-\varepsilon_3}{2\Delta}$ with $\varepsilon_1 > 0, \varepsilon_2 > 0$ & $\varepsilon_3 < \varepsilon_4$

Since $\|x - \hat{x}^{(k)}\| \leq \|x - \hat{x}^{(l)}\|$ and using (1) we have,

$\frac{\varepsilon_4-\varepsilon_3}{2\Delta} \leq \frac{\varepsilon_2-\varepsilon_1}{2\Delta} \implies \mathscr{G}^{(k)}(j) \leq \mathscr{G}^{(l)}(j)$

Proposition 1 *If* $x = (x_1, \ldots x_j \ldots x_n) \in \mathscr{M}$ *then the gradient defined by* (5) *will be* $\mathscr{G}^{(k)} = (0,0,\ldots 0)$.

$\mathscr{G}^{(k)}(j) = \frac{\left\|\hat{x}_{+\delta}-\mathscr{P}_{\mathscr{M}}(\hat{x}_{+\delta})\right\|_2 - \left\|\hat{x}_{-\delta}-\mathscr{P}_{\mathscr{M}}(\hat{x}_{-\delta})\right\|_2}{2\Delta}$

where $j \in \{1,2,\ldots n\}$

When $x \in \mathscr{M}$, *using the definition* (1) *we have the distances*

$\left\|\hat{x}_{+\delta} - \mathscr{P}_{\mathscr{M}}(\hat{x}_{+\delta})\right\|_2 = \left\|\hat{x}_{-\delta} - \mathscr{P}_{\mathscr{M}}(\hat{x}_{-\delta})\right\|_2$

$\therefore \mathscr{G}^{(k)} = (0,0,\ldots 0)$.

According *Theorem-1* stability can be achieved, if we choose parameter Δ such a way that $\frac{\varepsilon_2-\varepsilon_1}{2\Delta} \leq 1$. Therefore, proper selection of the Δ will lead to a stable recovery. It is clear from *Proposition-1* that when the signal $\hat{x}$ approaches clean signal x then the gradient $\mathscr{G}$ approaches zero. Also each estimate update $\hat{x}^{(k+1)} \leftarrow \hat{x}^{(k)} - \mu\mathscr{G}^{(k)}$ is performed in such a way that $\|x - \hat{x}^{(k+1)}\|_2 \leq \|x - \hat{x}^{(k)}\|_2$. Hence the proper selection of Δ and μ also ensures the convergence of the algorithm.

Sensitivity of hyper-parametes Δ and μ:

Detailed analysis of hyper-parameters Δ and μ is shown in the figure: 1, 2, 3 and 4 for translated Gaussian pulse recovery. When the analysis is carried out by varying μ, the other parameter is kept at $\Delta = 0.5$. Similarly when the analysis is carried out by varying Δ, the other parameter is kept at $\mu = 0.5$. Analysis reveals that increase in μ value result in early convergence. Better recovery also can be achieved by increasing value of μ. On the other hand, when μ is increased beyond unity the algorithm diverges. Algorithms goes unstable when the value μ approaches unity. So it recommended to limit the value of μ between zero and unity $(0 < \mu < 1)$ for better performance. Parameter Δ is application specific and shows lesser sensitivity compared to μ. It is oblivious from the analysis that the increase in Δ guarantee an early recovery and better performance, even if its value increases beyond unity. But the algorithm goes unstable, if it is increased beyond a certain limit. The limit depends on the current value of the signal at the position under consideration. Therefore, it is recommended to limit the value $0 < \Delta < 1$ for a stable recovery.

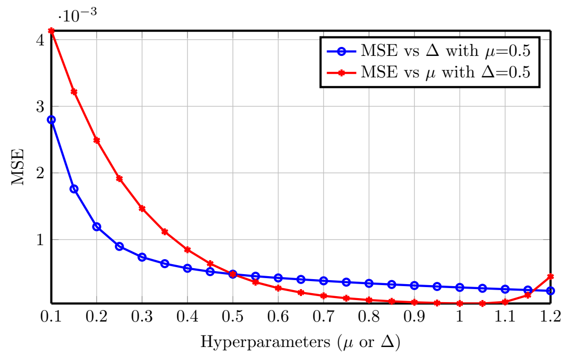

Fig. 1 Sensitivity of hyper-parameters: Dependence of MSE on μ *or* Δ with other parameter fixed at 0.5

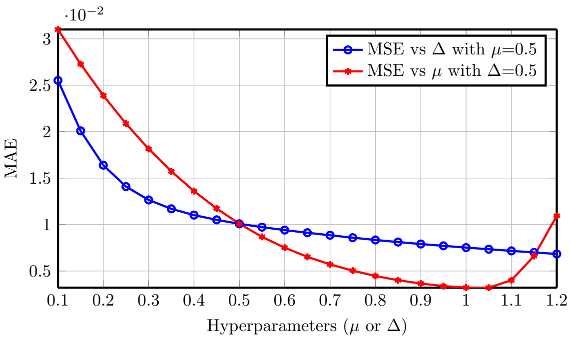

Fig. 2 Sensitivity of hyper-parameters: Dependence of MAE on μ *or* Δ with other parameter fixed at 0.5

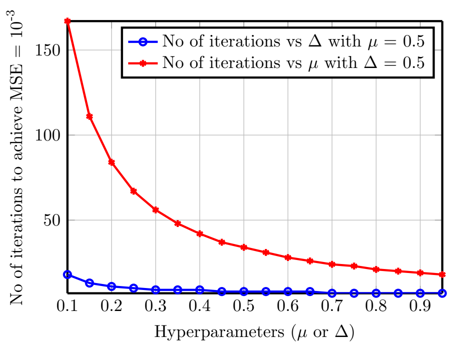

Fig. 3 Sensitivity of hyper-parameters: Number of iterations to reach a MSE = 10^{-3} vs μ *or* Δ with other parameter fixed at 0.65

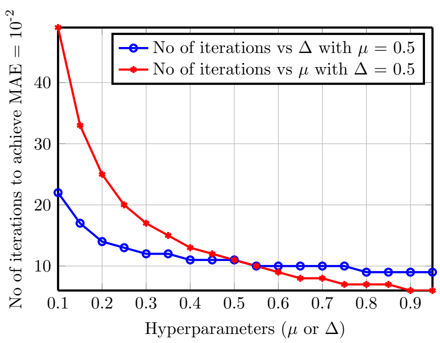

Fig. 4 Sensitivity of hyper-parameters: Number of iterations to reach a MAE = 10^{-2} vs μ *or* Δ with other parameter fixed at 0.5