基于压缩感知的振动数据修复方法*

张新鹏 胡茑庆† 程哲 钟华

(国防科学技术大学机电工程与自动化学院, 装备综合保障技术重点实验室, 长沙 410073)

(2014 年 3 月 27 日收到; 2014 年 6 月 9 日收到修改稿)

为解决旋转机械振动信号丢失数据修复的问题, 提出一种基于压缩感知原理的振动数据修复方法. 首先对采集到的不完整信号进行处理, 将无信息输入时刻对应的数据用零元素填充得到有损信号, 以单位矩阵为基础, 根据有损信号中零元素的位置信息, 构造对应于压缩感知框架下的观测矩阵; 再根据待修复信号的特点并结合先验知识, 构造或选择能够对振动信号进行稀疏表示的字典矩阵; 然后使用高效且稳定的追踪算法, 根据有损信号、观测矩阵以及字典矩阵重构原始信号, 实现丢失振动数据的修复. 使用仿真数据检验方法的有效性; 使用实测的轴承振动状态数据验证方法对于振动数据的适用性, 并通过比较完整信号、有损信号和修复信号对应的时域和频域特征值来检验数据修复效果. 实验结果表明: 本文方法能够有效地实现丢失数据的修复, 且从统计特征的角度来看, 相比于有损信号, 修复信号能够更为准确地描述真实完整的振动信号.

关键词: 数据修复, 压缩感知, 振动信号, 观测矩阵

PACS: 05.45.Tp **DOI:** 10.7498/aps.63.200506

1 引 言

在对旋转机械进行状态监控时, 常常需要采集设备关键部件如齿轮、轴承等运行过程中的振动数据, 并基于这些数据, 在时域、频域或者时频域进行故障的监测和诊断等. 由于监测环境常常伴随着较为强烈的振动, 有时会对传感器的正常工作产生一定的影响[1−4]. 如在实际振动信号采集过程中, 可能由于传感器故障、信号传输或者线路接触不良等原因, 信号采集过程突然中断, 导致部分时间段内的振动状态数据丢失. 在运行过程不可逆的情况下, 重新采样已不可能, 此时如需进行设备状态分析, 则不得不使用不完整数据, 而直接使用这些数据会影响故障检测或诊断结果的准确性. 如果能够根据实际采集得到的状态数据估计出采样中断期间所损失的数据, 则会提高设备状态分析结果的准确性.

压缩感知 (compressed sensing, CS) 是 Donoho[5], Candes 等[6,7] 于 2006 年正式提出的一种可突破 Nyquist 采样频率的新型采样理论[8,9], 其核心思想是将压缩过程和采样过程合并, 直接采集原始信号的少量非自适应线性投影 (测量值), 然后根据相应的重构算法由测量值重构得到相应的原始信号[10−12]. CS 出现之后, 便开始被研究和应用于诸多领域. 如在压缩成像领域, Chan 等[13] 研究了基于压缩感知的单像素太赫兹成像系统; 在关联成像领域, Katz 等[14] 将 CS 引入基于热光源的关联成像系统中, 白旭等[8] 则将 CS 理论与 DGI 相结合, 通过 CS 有效降低采样次数; 在医疗成像领域, Lustig 等[15] 将 CS 应用于核磁共振中, 实现快速成像; Herman 等[16] 利用 CS 实现了高分辨率雷达探测; 另外在无线通信[17]、天文[18]、模式识别[19] 及生物传感等领域也都开展了 CS 的应用研究工作.

对于丢失数据修复问题, 如果将当前采集到的不完整数据作为 CS 理论中的观测数据, 相应的完整数据作为原始信号, 则该数据修复问题便可等效为一个压缩采样过程. 实际上, 在 CS 原理提出之前, 已有文献利用相关的基础理论开展了图像修复方面的研究工作, 如文献 [20] 提出了一种利用 l_1 范

∗ 国家自然科学基金 (批准号: 51375484, 51205401) 资助的课题.

† 通讯作者. E-mail: huniaoqing@126.com

数最小化实现有损图像修复的方法; 在文献[21]中, Elad等提出了一种基于图像稀疏表示的图像修复算法, 并通过实验验证了算法的有效性; 文献[22]提出了一种利用线性规划方法解决丢失数据恢复的问题, 并分析了该方法实现数据恢复的条件. 上述涉及的范数最小化问题以及稀疏表示问题均可以包含在CS理论体系之内[23]. 本文在上述研究的基础上, 将CS原理引入丢失数据修复问题中, 并结合前文所述的应用背景, 提出了一种基于CS的振动数据修复方法. 该方法首先根据不完整数据构造CS框架下的观测矩阵, 然后针对振动信号的特点, 建立稀疏表示字典矩阵, 在此基础上利用追踪算法估计出完整的振动信号.

对于丢失数据的修复, 可用的方法还有时间序列法、回归分析法、插值法和曲线拟合法等. 这些方法是利用已有的数据信息提取出数据的分布规律或趋势, 根据这些规律来估计缺失的数据, 往往需要求解一系列的方程或高维方程组, 随着数据量的增大, 求解过程会越来越困难, 而且这些方法都是基于数据本身实现缺损数据修复, 并未考虑数据类型的特点. 本文所提方法不对数据趋势和规律进行显式提取, 这些信息是蕴含于数据重构过程中的, 从而将复杂的方程求解等问题转化为直接的矩阵运算. 同时, 本文方法在使用稀疏矩阵的过程中便将振动信号的类型特点考虑在内, 因此对数据的修复更为有利.

本文第二节介绍基于压缩感知的振动数据修复原理; 第三节通过仿真数据和实测轴承振动数据对本文所提方法的有效性进行验证; 最后在第四节对全文内容和下一步研究方向进行总结.

2 数据修复原理

根据CS原理, 假设可通过某种字典矩阵$\boldsymbol{\varPsi}$对原始信号$\boldsymbol{x} \in \mathbb{R}^N$进行稀疏化, 且经过压缩观测可得到观测信号$\boldsymbol{y} \in \mathbb{R}^M$ $(M < N)$. 此处的M个数据一般是通过随机观测得到的, 故无法确定M个数据中哪些点是重要的, 哪些点是不重要的. 由CS原理, 可以认为这M个点以很大的概率包含了原始信号$\boldsymbol{x}$的大部分重要信息. 而从与此相对的另一个角度来看, 也可以认为随机丢弃了$N-M$个数据, 且这$N-M$个数据里包含的原始信号$\boldsymbol{x}$的信息很少.

假设在原始信号$\boldsymbol{x} \in \mathbb{R}^N$的采集过程中, 由于诸如传感器接触不良等原因只得到了M个数据, 即丢失了$N-M$个数据. 这些数据的损失是有随机性的. 结合CS原理, 该过程可看作为一个随机压缩采样过程. 即由于丢失时刻的随机性, 可认为采集到的M个数据里包含了原始信号的大部分信息. 基于得到的这M个数据, 应用CS重构算法, 可以重构出原始信号N个数据的估计值. 至此实现了数据的修复. 上述数据修复思路的数学原理如下.

假设完整的原始信号可表示为$\boldsymbol{x} \in \mathbb{R}^N$, 采集得到的有损信号为$\boldsymbol{x}' \in \mathbb{R}^N$. 在采集程序中设定: 如若当前时刻未有数据输入, 则将该时刻对应数据值置为零(实际上, 在设备运行过程中, 振动传感器正常工作时输出值一般不等于零, 即使为零也可认为是随机情况, 故该设定不会对恢复算法造成影响). 假设$\boldsymbol{x}'$中包含M个非零元素, 则零值元素的个数为$N-M$. 设所有零值元素对应位置索引的集合为$\boldsymbol{Z}$, 即$\boldsymbol{Z} = \{i|x'_i = 0\}$ $(i = 1, 2, \cdots, N)$, 将$\boldsymbol{x}'$中$N-M$个零值去掉, 且保持其他元素值的顺序和大小不变, 可得到观测信号$\boldsymbol{y} \in \mathbb{R}^M$. 设$\boldsymbol{I}_n \in \mathbb{R}^{N\times N}$为单位矩阵, 将$\boldsymbol{I}_N$中的第$j$行$(j \in \boldsymbol{Z})$删除, 其他行保持顺序和大小不变, 可得到矩阵$\boldsymbol{\varPhi} \in \mathbb{R}^{M\times N}$. 该矩阵$\boldsymbol{\varPhi}$即相当于CS理论中的观测矩阵. 信号$\boldsymbol{y}$和矩阵$\boldsymbol{\varPhi}$的构造过程如图1所示, 其中白色方块对应为零值, 黑色方块对应值为1, 其他均为非零值.

根据上述定义有

$$\boldsymbol{y} = \boldsymbol{\varPhi}\boldsymbol{x}. \tag{1}$$

假设原始振动信号$\boldsymbol{x}$可以通过字典矩阵$\boldsymbol{\varPsi} \in \mathbb{R}^{N\times N}$进行稀疏化[24], 即

$$\boldsymbol{\varTheta} = \boldsymbol{\varPsi} \cdot \boldsymbol{x}, \tag{2}$$

其中$\boldsymbol{\varTheta}$为稀疏表示系数向量. 由(2)式可得

$$\boldsymbol{x} = \boldsymbol{\varPsi}^{-1} \cdot \boldsymbol{\varTheta}. \tag{3}$$

将(3)式代入(1)式, 并令$\boldsymbol{A} = \boldsymbol{\varPhi}\boldsymbol{\varPsi}^{-1}$, 则有

$$\boldsymbol{y} = \boldsymbol{\varPhi}\boldsymbol{\varPsi}^{-1}\boldsymbol{\varTheta} = \boldsymbol{A}\boldsymbol{\varTheta}. \tag{4}$$

此时已知观测值$\boldsymbol{y}$和测量矩阵$\boldsymbol{A}$, 需要重构$\boldsymbol{\varTheta}$, 而$\boldsymbol{\varTheta}$是稀疏的, 这便成为一个CS框架下的信号重构过程. Candés和Tao[12,25]指出并证明了重构需要满足的条件: $\boldsymbol{A}$必须满足约束等距性质(restricted isometry property, RIP). RIP的定义如下: 对任意k-稀疏的信号$\boldsymbol{\varTheta}$和常数$\delta_k \in (0, 1)$, 如果满足:

$$\begin{aligned}&(1-\delta_k)\|\boldsymbol{\varTheta}\|_2^2\\ &\leqslant \|\boldsymbol{A}\boldsymbol{\varTheta}\|_2^2 \leqslant (1+\delta_k)\|\boldsymbol{\varTheta}\|_2^2,\end{aligned} \tag{5}$$

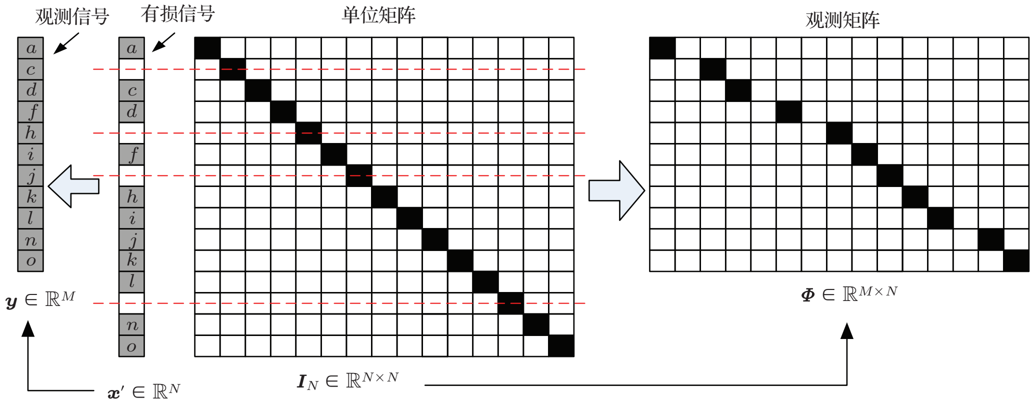

图1 信号 $\boldsymbol{y}$ 和矩阵 $\boldsymbol{\Phi}$ 的构造示意图

则称矩阵 $\boldsymbol{A}$ 满足RIP. 判断一个矩阵是否符合RIP本身就是一个组合复杂度问题. k-RIP只能作为一个充分条件, 用来判断测量矩阵是否满足特性, 但不是必要条件, 不能指导构造满足特性的测量矩阵, 因此单纯的提出 k-RIP依然无法实现测量矩阵的构造与实际使用. Baraniuk[26] 指出约束等距性质的等价条件是观测矩阵 $\boldsymbol{\Phi}$ 和稀疏表示字典矩阵 $\boldsymbol{\Psi}$ 不相关, 即要求 $\boldsymbol{\Phi}$ 的行 $\boldsymbol{\phi}_i^{\mathrm{T}}$ 不能由 $\boldsymbol{\Psi}$ 的列 $\boldsymbol{\psi}_j$ 稀疏表示, 且 $\boldsymbol{\Psi}$ 的列 $\boldsymbol{\psi}_j$ 不能由 $\boldsymbol{\Phi}$ 的行 $\boldsymbol{\phi}_i^{\mathrm{T}}$ 稀疏表示. 矩阵 $\boldsymbol{\Phi}\in\mathbb{R}^{M\times N}$ 和 $\boldsymbol{\Psi}\in\mathbb{R}^{N\times N}$ 的相关性定义为

$$\mu(\boldsymbol{\Phi},\boldsymbol{\Psi})=\sqrt{N}\cdot\max|\boldsymbol{\phi}_i^{\mathrm{T}}\boldsymbol{\psi}_j| \quad (1\leqslant i\leqslant M, 1\leqslant j\leqslant N), \tag{6}$$

其中 $\boldsymbol{\phi}_i^{\mathrm{T}}$ 和 $\boldsymbol{\psi}_j$ 分别表示矩阵 $\boldsymbol{\Phi}$ 和 $\boldsymbol{\Psi}$ 的第 i 行和第 j 列. 矩阵之间相关性越强, 互相表示时所需要的向量就越少. 为了确保原始信号 $\boldsymbol{x}$ 能够得到恢复, 有学者证明了稀疏度 k 和观测数量 M 之间须满足一定的关系[12,27,28]. 因此, 若 $\boldsymbol{A}$ 满足RIP, 或者矩阵 $\boldsymbol{\Phi}$ 和矩阵 $\boldsymbol{\Psi}$ 不相关, 则可重构得到稀疏向量 $\boldsymbol{\Theta}$ 的估计值 $\tilde{\boldsymbol{\Theta}}$[12,25,26]. 在得到 $\tilde{\boldsymbol{\Theta}}$ 后, 进一步可以计算完整信号 $\boldsymbol{x}$ 的估计 $\tilde{\boldsymbol{x}}$ 值, 即

$$\tilde{\boldsymbol{x}}=\boldsymbol{\Psi}^{-1}\cdot\tilde{\boldsymbol{\Theta}}, \tag{7}$$

至此实现了有损信号 $\boldsymbol{x}'$ 的修复. 根据CS原理, 本文所提方法的数据修复效果主要与矩阵 $\boldsymbol{\Phi}$ 的构建、稀疏字典矩阵 $\boldsymbol{\Psi}$(或待恢复信号在字典矩阵 $\boldsymbol{\Psi}$ 下的稀疏度 k) 以及所使用的重构算法有关. 在实际使用中应尽可能地选择或者构造能够使得原始信号足够稀疏的字典矩阵 $\boldsymbol{\Psi}$, 且确保 $\boldsymbol{\Psi}$ 与 $\boldsymbol{\Phi}$ 不相关. 数据重构时应尽可能使用恢复效果稳定且高效的重构算法[29,30]. 在上述修复方法中, 字典矩阵 $\boldsymbol{\Psi}$ 的选择不是惟一的, 能够将原始信号稀疏化的字典矩阵均可用于该修复方法中. 由于振动数据经过离散余弦变换 (discrete cosine transform, DCT) 后对应的系数是比较稀疏的, 因此下文中将以DCT矩阵为例, 利用仿真数据和实验数据对文中所提数据修复方法进行验证.

3 实验验证

前面介绍了基于CS的丢失信号恢复算法, 下面首先使用仿真数据检验方法的有效性, 然后使用实测的轴承振动状态数据来验证所提数据修复方法对振动数据的修复能力, 并通过比较原始信号和恢复信号对应的时域和频域特征值来检验数据修复效果.

3.1 仿真数据验证

首先构造仿真信号 $\boldsymbol{x}\in\mathbb{R}^N$, $N=1000$, 且 $\boldsymbol{x}$ 可以通过离散余弦变换进行稀疏化. DCT矩阵可表示为 $\boldsymbol{C}\in\mathbb{R}^{N\times N}$, 且

$$\boldsymbol{\Theta}=\boldsymbol{C}\cdot\boldsymbol{x}. \tag{8}$$

设置 $\boldsymbol{\Theta}$ 的稀疏度为 $k=80$, 即 $\boldsymbol{\Theta}$ 中非零元素的个数为80. 矩阵 $\boldsymbol{C}$ 的第 i 行第 j 列元素可通过 (9) 式进行计算[22]:

$$\boldsymbol{C}_{ij}=\frac{\min\{i,\sqrt{2}\}}{\sqrt{N}}\cos\left(\frac{(2j-1)(i-1)\pi}{2N}\right) \quad (1\leqslant i,j\leqslant N), \tag{9}$$

原始信号 $\boldsymbol{x}$ 及其在DCT矩阵下的稀疏表示系数向量 $\boldsymbol{\Theta}$ 如图2所示.

假定在数据采集过程中, 由于某种原因造成了部分数据丢失. 按照第三节所述方法, 利用得到的有损数据可以构造出对应的观测矩阵 $\boldsymbol{\Phi}$. 根据CS原理, 若要从有损数据中重构出完整数据, 需保证

观测矩阵 $\boldsymbol{\Phi}$ 和稀疏字典矩阵 $\boldsymbol{C}$ 不相关. 根据 (6) 式计算可得

$$\mu(\boldsymbol{I}_N, \boldsymbol{C}) = \sqrt{2}, \tag{10}$$

其中 $\boldsymbol{I}_N \in \mathbb{R}^{N\times N}$ 为单位矩阵. 由观测矩阵 $\boldsymbol{\Phi}$ 的构造过程可知

$$\mu(\boldsymbol{\Phi}, \boldsymbol{C}) \leqslant \sqrt{2}. \tag{11}$$

根据相关性的计算公式 (6), 对于任意两个 $N \times N$ 维的正交矩阵 $\boldsymbol{\Phi}$ 和 $\boldsymbol{\Psi}$ 而言, 二者相关性 $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ 的取值范围为 $[1, \sqrt{N}]$, 若取 $N = 1000$, 则理论上观测矩阵 $\boldsymbol{\Phi}$ 和稀疏矩阵 $\boldsymbol{\Psi}$ 的相关性的取值范围为 $[1, \sqrt{1000}]$, 而此处 $\boldsymbol{\Phi}$ 和 $\boldsymbol{C}$ 的相关性小于 $\sqrt{2}$, 因此可判断观测矩阵和稀疏矩阵具有较强的不相关性, 这意味着原始信号的重构是可能的. 因此在仿真实验中, 我们选择 DCT 矩阵作为本文所提方法中的字典矩阵. 在信号重构时使用正交匹配追踪 (orthogonal matching pursuit, OMP) 算法 [31−33].

设采样得到的数据个数 $M = 600$, 则丢失数据个数为 $N - M = 400$. 应用本文所提方法得到的数据修复结果如图 3 所示. 从图 3 中可以看出, 1000 点的原始数据在随机损失 40% (即 400 个点) 时仍可近乎完美的恢复出原始数据.

下面分析随机损失不同数据量时, 本文方法的修复效果. 数据修复效果以平均相对误差 $\bar{e}$ 来衡量, $\bar{e}$ 可通过下式计算得到

$$\bar{e} = \frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{x}_i - x_i}{x_i}, \tag{12}$$

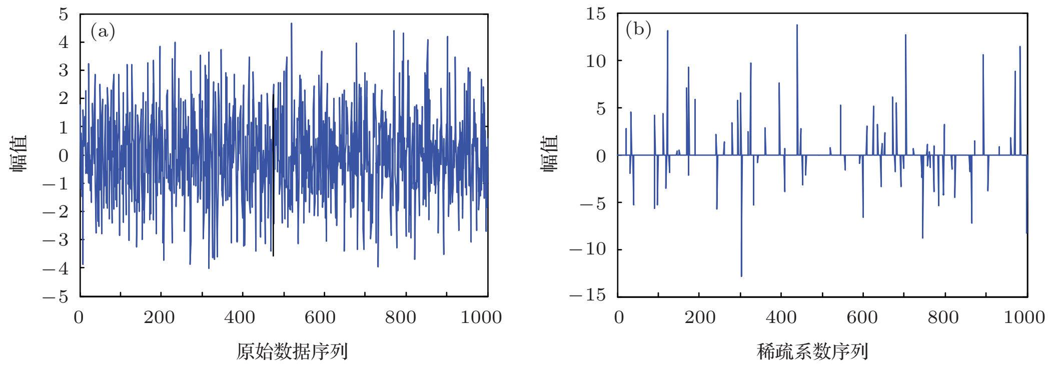

图 2 仿真信号 $\boldsymbol{x}$ (a) 及其对应于 DCT 矩阵的稀疏系数向量 $\boldsymbol{\Theta}$ (b)

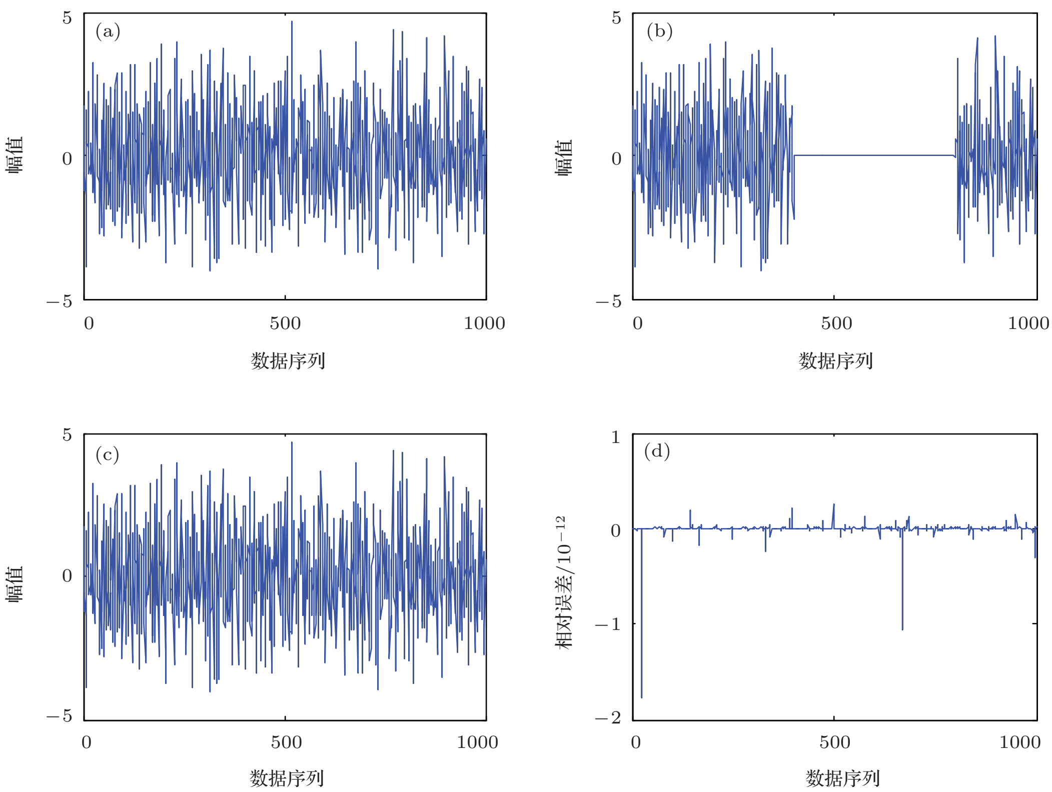

图 3 损失 400 点时的仿真数据修复结果

其中 $\tilde{x}_i$ 为恢复出的第 i 个元素, x_i 为原始信号 $\boldsymbol{x}$ 的第 i 个元素. 在保持恢复算法中相关参数不变的情况下, 不同数据损失量对应的数据恢复结果如图 4 所示. 从图 4 中可以看出, 当损失比低于 45% 时 (即损失量少于 $45\% \times 1000 = 450$ 个数据), 本文所提方法均可以近乎完美地恢复出原始数据. 损失数据量大于 45% 时平均相对误差很大, 也即无法恢复出原始的完整数据. 实际采集得到的数据量越多, 意味着数据损失量越少, 实现数据修复的可能性也就越大.

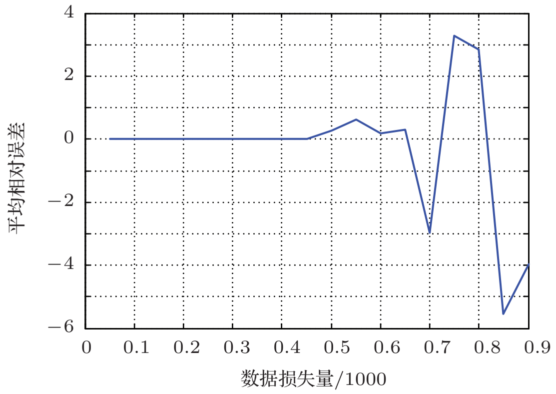

图 4　不同数据损失量对应的信号修复平均相对误差

3.2　实测振动数据验证

上一节通过仿真数据验证了本文所提数据修复方法的有效性, 下面使用实测时域振动信号来验证算法的数据修复能力. 设 $\boldsymbol{x} \in \mathbb{R}^N$ $(N = 1000)$ 为实测正常轴承运行时的时域振动信号 (实验轴承为 6205-2RS JEK SKF 深沟球轴承, 采样频率 12 K, 转速 1797 r/min, 负载 1 HP)[34], 选择 DCT 矩阵 $\boldsymbol{C}$ 作为字典矩阵, 计算 $\boldsymbol{\Theta} = \boldsymbol{C} \cdot \boldsymbol{x}$, 原始信号 $\boldsymbol{x}$ 及其在 DCT 矩阵下的系数向量 $\boldsymbol{\Theta}$ 如图 5 所示.

对于上述振动信号 $\boldsymbol{x} \in \mathbb{R}^N$ $(N = 1000)$, 假设在实际采集过程中只采集到了 $M = 900$ 个数据, 即丢失了 $N - M = 100$ 个数据. 在信号重构时同样使用 OMP 算法, 则本文方法的数据修复结果如图 6 和图 7 所示. 图 6 为修复结果的时域表示, 图 7 表示修复结果的频域表示.

从图 6 和图 7 中可以看出, 此时并不能精确地修复损失数据的每一点, 但仍可恢复出损失信号的总体趋势. 实际中我们常常会对振动信号的时域特征和频域特征进行分析, 即通常我们更关注的是振动信号的总体统计特征而不是某一点的具体值. 为了衡量振动信号的修复效果, 分别从原始信号、采集到的有损信号以及修复信号中提取出若干时域特征和频域特征, 通过对比这些特征值来分析数据的修复效果. 良好的数据修复效果意味着修复结果可更为准确地描述原始信号 (相比直接使用有损信号). 对于振动信号, 常用的时域特征和频域特征如表 1 和表 2 所示 [35]. 这些特征描述了振动信号的统计特性.

从上述特征中选择若干个时域特征和频域特征, 选择的依据是原始信号和有损信号计算得到的这些特征差距明显. 利用先验知识, 模拟相同的丢失事件 (丢失位置和数量), 分别计算原始信号和有损信号对应的时域和频域特征值, 结果如表 3 所示.

结合上述先验知识, 从表 3 中选择数据损失后变化较大的特征. 最终选择的时域特征指标为 T_2, T_5, T_{10} 和 T_{12}, 频域指标为 F_3, F_4, F_5, F_6, F_7, F_8, F_{11}, F_{12}. 图 6 和图 7 中的信号对应的特征计算结果如表 4 所示.

从表 4 中可以看出, 当损失比为 10% 时 (即损失数据量为 100 点), 相比有损信号, 修复信号计算得到的上述特征值更为接近原始完整信号对应的特征值. 实际中计算振动信号特征值时 (不局限于文中提及的特征, 也可使用其他能够有效描述信号统计规律的特征), 可使用修复信号来计算所选特征, 其他特征则使用有损信号计算. 这样可保证所有特征值都能够尽可能地逼近原始信号特征值. 因

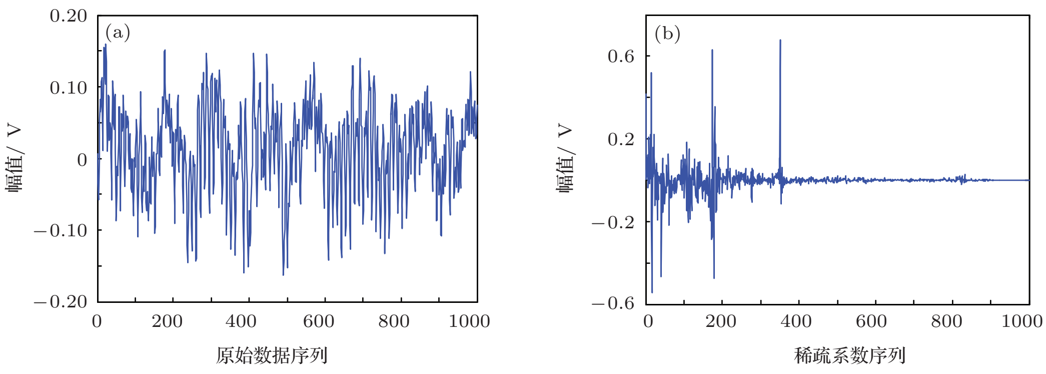

图 5　实测振动信号 $\boldsymbol{x}$ (a) 及其在 DCT 矩阵下的系数向量 $\boldsymbol{\Theta}$ (b)

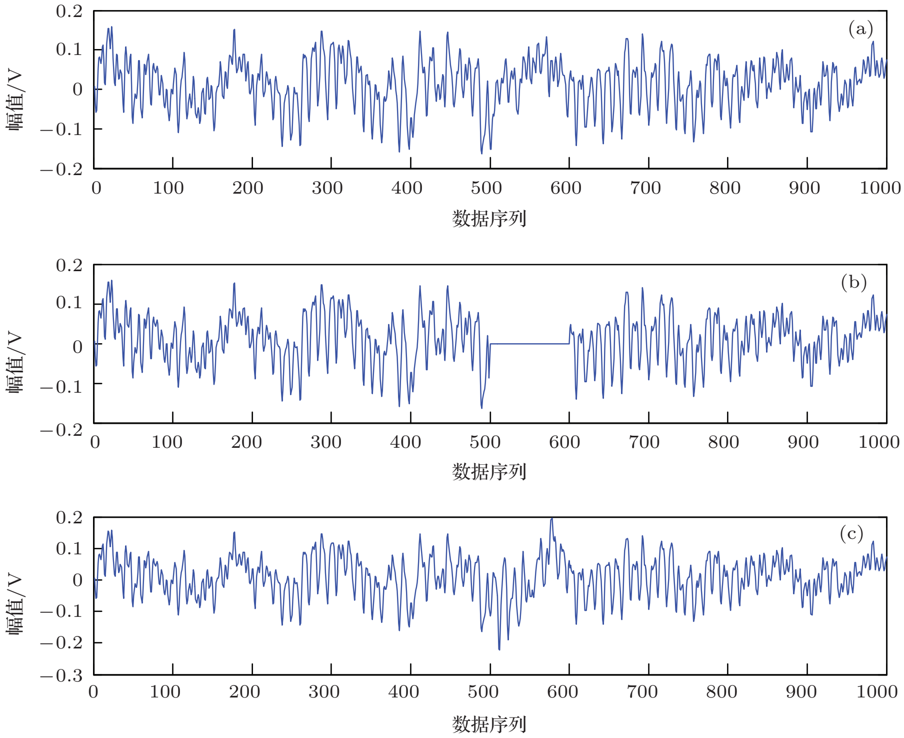

图6　损失100点数据时的振动数据修复结果(时域)　(a) 原始信号; (b) 有损信号(损失100点); (c) 修复信号

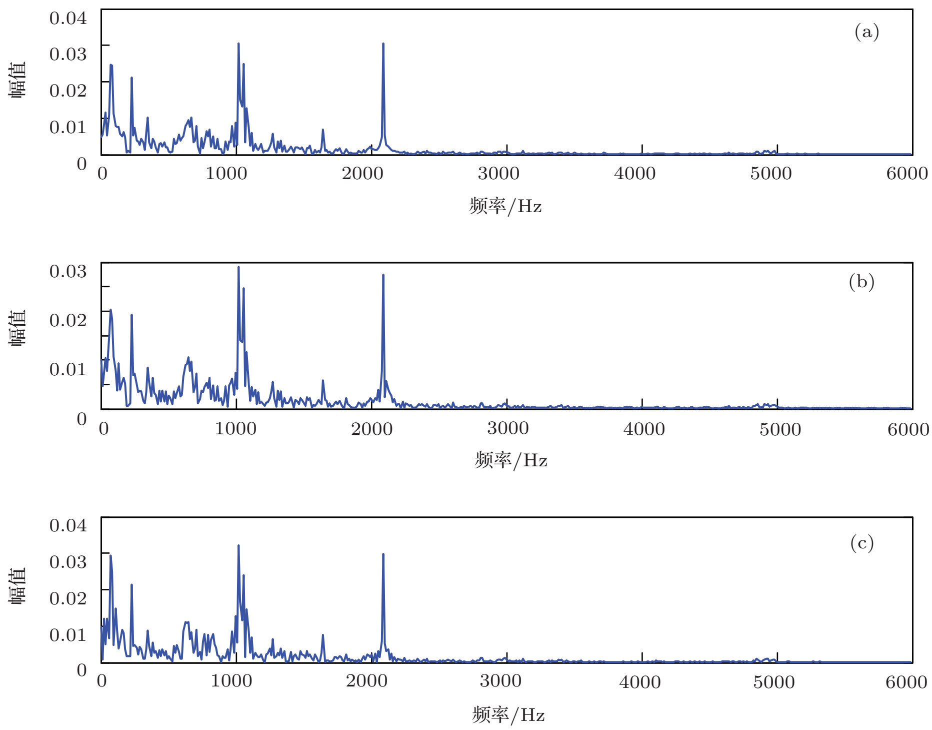

图7　损失100点数据时的振动数据修复结果(频域)　(a) 原始信号; (b) 有损信号(损失100点); (c) 修复信号

此, 从统计特征的角度来看, 利用本文数据修复方法可更为准确地描述原始信号, 从而证明了所提数据修复方法的有效性. 数据损失量为200点时, 分析过程与前述类似. 从特征值的变化来看, 此时数据修复效果相较于损失100时略有下降, 且随着损失数据的增多, 数据修复效果逐渐下降. 然而, 当损失点数过少时, 数据修复效果反而会变得不稳定, 这是因为此时对信号统计特征而言, 信号修复带来的影响有时要大于数据损失所带来的影响. 上述实验分析了轴承正常状态振动信号的数据修复效

表 1 时域统计特征

特征	T_1	T_2	T_3	T_4	T_5
计算方法	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2$	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	$\frac{1}{N}\sum_{i=1}^{N}(x_i - T_1)^2$	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$
特征	T_6	T_7	T_8	T_9	T_{10}
计算方法	$\max(\boldsymbol{x})$	$T_6 - \min(\boldsymbol{x})$	T_6/T_2	T_6/T_3	T_3/T_5
特征	T_{11}	T_{12}			
计算方法	T_6/T_5	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - T_1}{T_3}\right)^3$			

注: N 为数据总数, x 为样本序列, i 为数据索引.

表 2 频域统计特征

特征	F_1	F_2	F_3
计算方法	$\frac{1}{N}\sum_{i=1}^{N} y_i$	$\sum_{i=1}^{N}(y_i - F_1)^2$	$\frac{\sum_{i=1}^{N}(y_i - F_1)^3}{N \times \sqrt{F_2^3}}$
特征	F_4	F_5	F_6
计算方法	$\frac{\sum_{i=1}^{N}(y_i - F_1)^4}{N \times F_2^2}$	$\frac{\sum_{i=1}^{N} f_i y_i}{\sum_{i=1}^{N} y_i}$	$\sqrt{\frac{\sum_{i=1}^{N}((f_i - F_5)^2 \cdot y_i)}{N}}$
特征	F_7	F_8	F_9
计算方法	$\sqrt{\frac{\sum_{i=1}^{N}(f_i^2 \cdot y_i)}{\sum_{i=1}^{N} y_i}}$	$\sqrt{\frac{\sum_{i=1}^{N}(f_i^4 \cdot y_i)}{\sum_{i=1}^{N}(f_i^2 \cdot y_i)}}$	$\frac{\sum_{i=1}^{N}(f_i^2 \cdot y_i)}{\sqrt{\sum_{i=1}^{N} y_i \times \sum_{i=1}^{N}(f_i^4 \cdot y_i)}}$
特征	F_{10}	F_{11}	F_{12}
计算方法	F_6/F_5	$\frac{\sum_{i=1}^{N}((f_i - F_5)^3 \times y_i)}{N \times F_6^3}$	$\frac{\sum_{i=1}^{N}((f_i - F_5)^4 \times y_i)}{N \times F_6^4}$

注: N 为数据总数, y 为频域幅值序列, i 为数据索引, f 为频率序列.

表 3 原始信号、有损信号对应的特征值 (先验知识)

信号	时域特征											
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
原始信号	0.010	0.044	0.069	0.004	0.053	0.197	0.375	4.436	2.950	1.259	3.715	−0.35
有损信号	0.006	0.034	0.061	0.003	0.046	0.197	0.375	4.721	3.238	1.429	4.303	−0.26
信号	频域特征											
	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
原始信号	0	0	4.8	31.4	1027	42.2	1455	3333	0.4	0	43.6	4039
有损信号	0	0	4.2	25.5	1139	47.1	1611	3603	0.4	0	41.2	3633

表 4 原始信号、有损信号与修复信号对应的特征值

信号	特征					
	T_2	T_5	T_{10}	T_{12}	F_3	F_4
原始信号	0.0440	0.0510	1.2200	−0.3656	4.8	31.5
有损信号	0.0353	0.0458	1.2911	−0.2664	4.4	27.5
修复信号	0.0454	0.0530	1.2345	−0.3507	4.6	29.2
信号	特征					
	F_5	F_6	F_7	F_8	F_{11}	F_{12}
原始信号	1090.1	44.3	1544.5	3515.0	44.0	3983.1
有损信号	1172.1	47.8	1653.8	3686.5	42.0	3633.0
修复信号	1093.1	44.0	1526.6	3428.1	42.6	3857.5

果, 此外我们还对故障状态 (内环故障状态和外环故障状态) 对应的振动信号进行了数据修复分析, 修复效果与上述实验结论类似.

下面通过对比实验来验证文中所提方法的有效性. 应用插值法和曲线拟合法对图 6 所示振动信号丢失数据进行修复, 比较上述方法与本文方法的修复效果, 结果如图 8 所示. 其中图 8 (c) 为应用本文方法得到的修复结果; (d) 为应用三次多项式插值法得到的修复结果; (e) 为应用曲线拟合法得到的修复结果. 对比原始信号 (a) 可以直观地看出, 使用本文方法所得的数据修复结果优于传统的三次多项式插值和曲线拟合法, 从而验证了本文方法对振动数据修复的高效性.

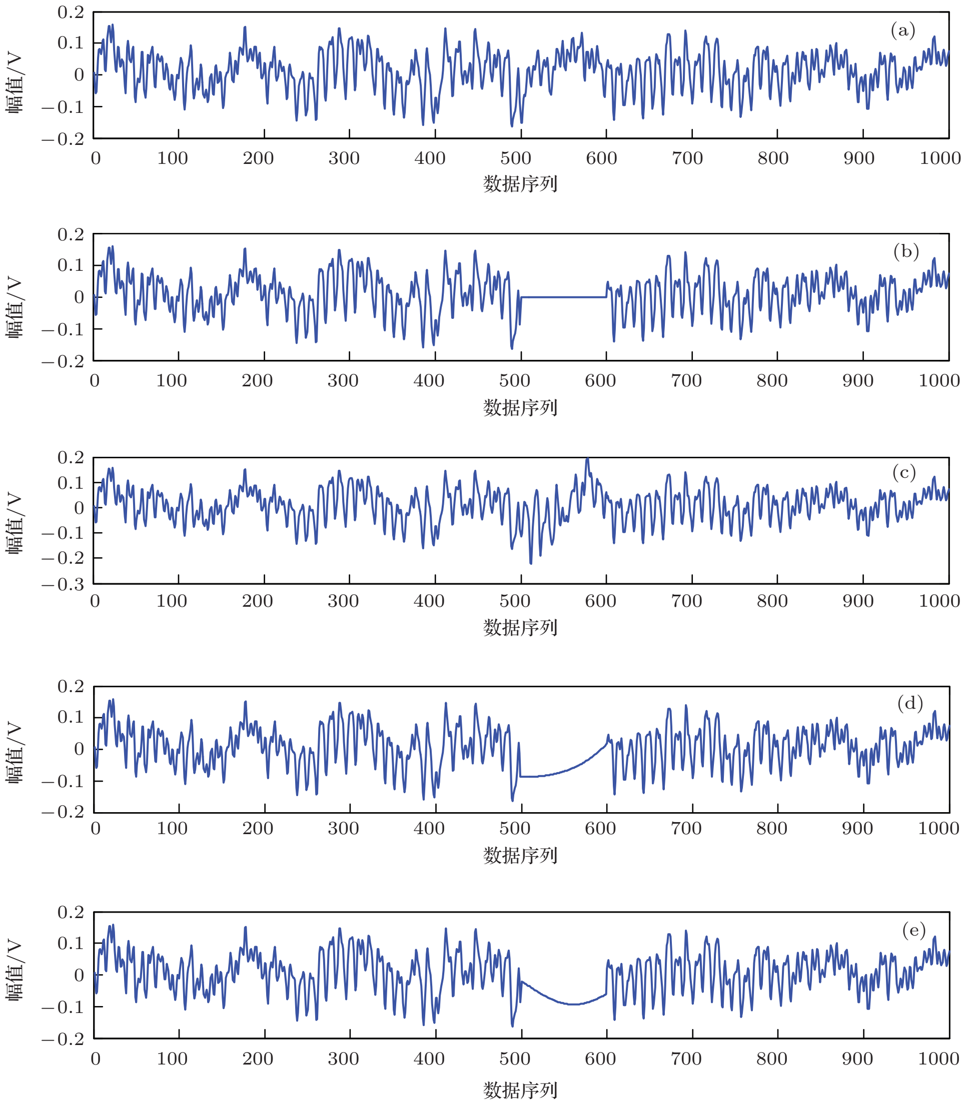

图 8　不同方法对应的振动数据修复效果　(a) 原始信号; (b) 有损信号 (损失 100 点); (c) 本文方法修复信号; (d) 三次多项式插值法修复结果; (e) 曲线拟合法 (阶数 20) 修复结果

4 结　论

本文提出了一种基于 CS 原理的振动信号修复方法, 并使用仿真数据和实测振动数据对所提方法进行了验证. 对于仿真信号, 因其可通过所使用的字典矩阵实现稀疏化, 故可得到近乎完美的数据修复结果. 对于实测振动信号, 由于噪声及信号传输等因素的影响, 无法在时域或者频域精确地恢复出每个丢失数据, 但是可明显地改善信号的统计特性, 而振动信号的统计特征往往是状态分析过程中所重点关注的. 实验结果表明, 从统计特征角度来看, 在振动数据损失量较少的情况下, 本文方法可以得到良好的数据修复效果, 能够更为准确地描述原始信号. 根据 CS 原理, 影响本文数据修复效果的主要因素包括稀疏字典以及重构算法. 文中针对

振动信号使用DCT矩阵作为字典矩阵, 实际上如果能够构造出使得振动信号更为稀疏的字典矩阵, 或者使用更为高效的重构算法, 将有利于修复效果的提高, 这也是下一步的研究方向. 此外, 从数据修复原理可以看出, 本文方法只要求待分析信号是稀疏的或者能够在某个字典矩阵下进行稀疏化, 因此也可用于解决其他可稀疏表示的一维信号如声学信号等的丢失数据修复问题.

参考文献

[1] Gu Y, Li Q, Xu B J, Zhao Z 2014 *Chin. Phys. B* **23** 017804

[2] Gao W, Zha F S, Song B Y, Li M T 2014 *Chin. Phys. B* **23** 010701

[3] Yu G Y, Song Y F, He X, Zheng X X, Tan D W, Chen J, Yang Y Q 2012 *Chin. Phys. B* **21** 043402

[4] Liang X X, Ban S L 2004 *Chin. Phys.* **13** 71

[5] Donoho D L 2006 *IEEE Trans. Inf. Theory* **52** 1289

[6] Candés E J, Romberg J, Tao T 2006 *IEEE Trans. Inf. Theory* **52** 489

[7] Candés E J 2006 *Proceedings of International Congress of Mathematicians* (Switzerland: European Mathematical Society Publishing House) p1433

[8] Bai X, Li Y Q, Zhao S M 2013 *Acta Phys. Sin.* **62** 044209 (in Chinese) [白旭, 李永强, 赵生妹 2013 物理学报 **62** 044209]

[9] Zhao S M, Zhuang P 2014 *Chin. Phys. B* **23** 054203

[10] Candés E J, Wakin M 2008 *IEEE Signal Proc. Mag.* **25** 21

[11] Donoho D L, Tsaig Y 2006 *Signal Process.* **86** 533

[12] Candés E J, Tao T 2006 *IEEE Trans. Inf. Theory* **52** 5406

[13] Chan W L, Charan K, Takhar D, Kelly K F, Baraniuk R G, Mittleman D M 2008 *Appl. Phys. Lett.* **93** 121105

[14] Katz O, Bromberg Y, Silberberg Y 2009 *Appl. Phys. Lett.* **95** 131110

[15] Lustig M, Donoho D, Pauly J M 2007 *Magn. Reson. Med.* **58** 1182

[16] Herman M, Strohmer T 2009 *IEEE Trans. Signal Process* **57** 2275

[17] Bajwa W U, Haupt J 2007 *IEEE Trans. Signal Process* **53** 3629

[18] Bobin J, Starck J L, Ottensamer R 2008 *Appl. Math. Comput.* **206** 980

[19] Elad M 2007 *IEEE Trans. Signal Process* **55** 5695

[20] Selesnick I, van Slyke R, Guleryuz O 2004 *International Conference on Image Processing (ICIP)* Singapore, October 24–27, 2004 p1819

[21] Elad M, Starck J L, Querre P, Donoho D L 2005 *Appl. Comput. Harmon A* **19** 340

[22] Zhang Y 2006 *Rice University CAAM Technical Report* Houston, USA, October, 2006 TR 06–15

[23] Ning F L, He B J, Wei J 2013 *Acta Phys. Sin.* **62** 174212 (in Chinese) [宁方立, 何壁静, 韦娟 2013 物理学报 **62** 174212]

[24] Feng B C, Fang S, Zhang L G, Li H, Tong J J, Li W Q 2013 *Acta Phys. Sin.* **62** 112901 (in Chinese) [冯丙辰, 方晟, 张立国, 李红, 童节娟, 李文茜 2013 物理学报 **62** 112901]

[25] Candés E J, Tao T 2005 *IEEE Trans. Inf. Theory* **51** 4203

[26] Baraniuk R 2007 *IEEE Signal Proc. Mag.* **24** 118

[27] Davenport M 2010 *Ph. D. Dissertation* (Houston: Rice University)

[28] Krahmer F, Ward R 2011 *SIAM J. Math. Anal.* **43** 1269

[29] Sun Y L, Tao J X 2014 *Chin. Phys. B* **23** 078703

[30] Wang L Y, Li L, Yan B, Jiang C S, Wang H Y, Bao S L 2010 *Chin. Phys. B* **19** 088106

[31] Pati Y C, Rezaiifar R, Krishnaprasad P S 1993 *Twenty-seventh Asilomar Conference on Signal, System & Computers* Pacific Grove, California, November 1–3, 1993 44

[32] Mallat S, Davis G, Zhang Z 1994 *SPIE J. Opt. Eng.* **33** 2183

[33] DeVore R A, Temlyakov V N 1996 *Adv. Comput. Math.* **5** 173

[34] http://csegroups.case.edu/bearingdatacenter/pages/download-data-file

[35] Sun C, He Z J, Zhang Z S, Chen X F, Cao H R, Ning X Y, Zou L M 2013 *Chin. J. Mech. Engineer.* **49** 30 (in Chinese) [孙闯, 何正嘉, 张周锁, 陈雪峰, 曹宏瑞, 宁喜钰, 邹利民 2013 机械工程学报 **49** 30]

Vibration data recovery based on compressed sensing*

Zhang Xin-Peng Hu Niao-Qing† Cheng Zhe Zhong Hua

(*Science and Technology on Integrated Logistics Support Laboratory, College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China*)

(Received 27 March 2014; revised manuscript received 9 June 2014)

Abstract

A missing data recovery method based on compressed sensing is proposed for vibration data of rotating machinery. Firstly, the incomplete signal is transformed into lossy signal by setting the data values corresponding to the time without input as zeros. According to the indices of zero elements in lossy signal, the observation matrix in the frame of compressed sensing is constructed based on identity matrix. Secondly, the dictionary matrix with which the vibration signal can be represented sparsely is chosen or constructed according to the signal needed to be recovered and other prior knowledge. Finally, the original complete signal is recovered based on the lossy signal, observation matrix and dictionary matrix by using an effective and steady pursuit algorithm. The efficiency of the proposed method is validated with simulation data and practical bearing vibration data. Recovery results are discussed by comparing the characteristic values corresponding to the complete signal, lossy signal and recovered signal in time domain and frequency domain. The test results show that the proposed method can well achieve the missing data recovery, and from the view of statistical characteristics, the recovery signal can describe the complete vibration signal more accurately than the lossy signal.

Keywords: data recovery, compressed sensing, vibration signal, observation matrix

PACS: 05.45.Tp **DOI:** 10.7498/aps.63.200506

* Project supported by the National Natural Science Foundation of China (Grant Nos. 51375484, 51205401).

† Corresponding author. E-mail: huniaoqing@126.com