Fault Diagnosis, Prognosis and Reliability of Electrical Drives

Elias Strangas and Selin Aviyente

Department of Electrical and Computer Engineering
Michigan State University
strangas@egr.msu.edu, aviyente@egr.msu.edu
Overview

1. Part I: Objectives of Failure Prognosis
2. Part II: Fault Diagnosis – Methods and Examples
3. Part III: Failure Prognosis – Methods and Examples
4. Part IV: Reliability and Directions in Research
Objectives of Fault Diagnosis

At the basic level

- Detect abnormal operation of a subsystem or system,
- Determine which component is failing,
- Estimate how it is failing, and how severe the fault is.

Next steps

- Evaluate the information of the type, severity and confidence of the fault determination,
- Schedule maintenance, based on fault severity and operating requirements and conditions,
- Alternatively, employ redundancies.
Fault diagnosis

Diagnosis of a non-catastrophic fault requires:

- a data or a physics based model, based on the fault characteristics, or alternatively *a priori* training, based on observations of known faults.

- a method to extract a limited number of features from observations,

- a classification, i.e. a signal processing method to make determinations from these.
What faults to expect in Electrical Drives

- Bearing faults: affected by wear, temperature, loading, environment,
- Insulation: temperature, overvoltage, initial manufacturing quality,
- Connections: welding, crimping, corrosion,
- Rotor eccentricity: manufacturing, loading, wear,
- Rotor bar breakage in induction motors: manufacturing problems, starting cycles,
- Permanent magnet demagnetization: load, temperature, controller error, noise
- Gears,
- Sensor failure (e.g. rotor position sensor, current sensor).
- Power electronics components: switches, capacitors, gate drivers.
Fault tolerance and built-in redundancies

Addressing winding short circuit

- Single-layer fractional-slot windings
- High phase inductance
- High number of phases
- Control algorithm (short a phase – inject d-axis current)

Addressing winding or inverter open circuit

Corresponding increase in stator currents
Fault tolerance and built-in redundancies

Thermal management

- Decreased winding losses by decreasing currents – this in turn requires changes in torque
- Change of switching frequency for inverters. This affects losses in the conductors and junction temperatures of switches
- Reduced voltage to decrease iron losses with effects on speed, switching and DC link.
Determining faults and fault severity

Model-based techniques

- Require that a model of the device is available and running in parallel to the operation, or
- at least, have prepared anticipated performance,
- Sensors measure this performance, and the two are compared,
- A logic, part of the controller makes a decision

Examples

- Winding short or open – zero sequence voltages or currents, other deviations of currents
- broken rotor bars in induction motors: stator currents of frequency of $(1 \pm 2s)f$
- Frequencies associated with eccentricities
Determining faults and fault severity

Data-driven techniques

- Require *a priori* training of the algorithm.
  - a relatively large collection of samples with known faults of all states
  - extraction of features, testing of the categorization algorithm
  - extraction of statistics

- Steps:
  - Data collection, preprocessing,
  - feature extraction,
  - fault classification,
  - decision
Determining faults and fault severity

Sensors and characteristics

- They define the cost of fault diagnosis more than any other part,
- Preferred: sensors that are there already, typically low bandwidth phase current sensors, occasionally DC link and phase voltage sensors
- Accelerometers and microphones, for vibrations etc.

Data storage and processing

- They define the cost of fault diagnosis more than any other part,
- Data are collected in batches or “epochs” and are processed almost in real time,
- Stored are the features rather than the raw data.
Determining faults and fault severity

Feature Extraction Methods
- Short time Fourier transform
- Undecimated wavelet transform
- Wigner-Ville transform
- Choi-Williams transform

Diagnosis (Detection & Categorization)
- Linear discriminant classifier
- Nearest neighbor classifiers
- Multiple discriminant classifier
- Support vector machine classifier
Diagnostic Methods

**Linear Discriminant Classifier**
- Discriminant function
  \[ D_k(x) = x_1\alpha_{1k} + \ldots + x_{N+1}\alpha_{1+Nk} \]
- Categorization
  \[ D_j(x) > D_k(x) \quad \forall k \neq j \]

**Nearest Neighborhood Classifiers**
- Compute N dimensional centroids
- Categorization:
  - Euclidean Distance
    \[ D_j(x) = \sqrt{(C_j - X)^2} \quad C, X \in \mathbb{R}^N \]
  - Mahalanobis Distance
    \[ D_j(x) = \sqrt{(C_j - X)\Sigma_j^{-1}(C_j - X)^{-1}} \quad C, X \in \mathbb{R}^N \]

**Multiple discriminant classifier**
- Project data in lower dimensional space by the optimal projection matrix \( W \)
- Can use LDC and NNCs
- Highly sensitive to training data

**Support vector machine classifier**
- Projects data in higher dimensional space
- Separating planes are computed
- Performs categorization in 2 classes
Example - LDC

A sample features vector belongs to a particular class if its discriminant function for that class is greater than for the any other class, i.e., $x$ belongs to class $j$ if:

$$D_j(x) > D_k(x)$$

for every $k \neq j$.

The weighting coefficients are adjusted from their initial guess through a training procedure.

For multiclass problems ($K > 2$), the classes are linearly separable if linear discriminant functions $D_1(x), \ldots, D_K(x)$ exist, such that the following is true.

$$D_j(x) > D_k(x) \quad \text{for every } x \text{ in } C_j \text{ and all } k \neq j$$

(1)
Example - LDC

We applied the method to a DC motor with a gear box, where the gears had a varying number of damaged teeth. We used the DC current to detect the fault and Choi-Williams for feature extraction.

Five states were identified: healthy; fault severity 1; fault severity 2; fault severity 3 and fault severity 4.

The data used to train the algorithm were 127 CWD samples of the high frequency bands from 15 experiments for each severity level of fault. From each experiment, a total 300 samples (60 per class) were used for training. Twenty events from each of the operating conditions were tested.

<table>
<thead>
<tr>
<th></th>
<th>Tested</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>20</td>
<td>20</td>
<td>00</td>
</tr>
<tr>
<td>Severity 1</td>
<td>20</td>
<td>12</td>
<td>08</td>
</tr>
<tr>
<td>Severity 2</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Severity 3</td>
<td>20</td>
<td>07</td>
<td>13</td>
</tr>
<tr>
<td>Severity 4</td>
<td>20</td>
<td>12</td>
<td>08</td>
</tr>
</tbody>
</table>
Objectives of Failure Prognosis

At the basic level

• Should maintenance be performed at the next scheduled time, or earlier?
• Is the drive available for the next task?

Beyond this

• What is the Remaining Useful Life of the drive?
• How much should I trust that estimation?
• If there is a developing fault, what can be done to delay or avoid failure?
What is needed for Prognosis?

- First, some method to estimate the state of the component or subsystem and the associated probabilities.
- Second, if there is a fault some technique to evaluate the evolution of the fault
  – This can be a physics-based model with uncertainties,
  – or a data-based model.
- A method to use all these to determine the expected state of the fault in the next interval/sample
- Some threshold relating this expectation (and the confidence in it) to failure.
- Although not part of prognosis exactly, a plan of action (mitigation, redundancy, emergency shutdown, scheduled maintenance, …)
Failure prognosis requires both:

- Extensive test data, usually from observations of artificially created faults, and a statistical model resulting from them.
- A physical model that will predict fault progression.
Prognosis – Physical models for failure progression in drives

- Insulation: expected life deteriorates with temperature, Arrhenius model
- Bearings: measure debris and estimate spall size and propagation

\[ R(t) = Ae^{-\frac{E}{kT}} \]


Electronic switches can be monitored, and a relation established between thermal cycling and aging. Die attach damage is a main failure mechanism, drain to source on-resistance is a precursor of failure.

Celaya, Towards Prognostics of Power MOSFETs: Accelerated Aging, and Precursors of Failure, 2010
**Prognosis methods – Baysian Methods**

- Parameter space, $\Theta$, in our case all our possible states,
- Sample space, $Y$, possible datasets, in our case features of samples,
- *Prior distribution*: $p(\theta)$, our belief that $\theta$ represents the state,
- For each $\theta$ and $y$, a sampling model $p(y|\theta)$ our belief that $y$ will be the outcome if $\theta$ is true
- Update: for each value of $\theta$, a possible state, the *posterior* distribution $p(\theta|y)$ describes our belief that $\theta$ is the true value.

$$p(\theta|y) = \frac{p(\theta|y)p(\theta|y)}{\sum_{\theta}(y|\theta)p(\theta|y)}$$
Hidden Markov Models

- A statistical modeling method which assumes states to be Markovian.
  - A limitation, since history does not matter (e.g. fatigue)
  - hidden because we do not know the states

- States here correspond to discrete levels of fault severity.

- Finds the hidden variables (machine states) from the observable parameters (features extracted from sampled signals).
**Hidden Markov Models - Parameters**

*Problem*: Given the observation sequence \( y = \{y_1, y_2, \ldots, y_k\} \) and set of model parameters \( \theta = \{\pi, A, B\} \) how to choose corresponding state sequence \( x = \{x_1, x_2, \ldots, x_k\} \), which is optimal to generate the observation sequence. The optimal measure can be the maximum likelihood.

The model parameters are:

- State transition probability matrix \( (a_{ij} = p(x_{t+1} = j | x_t = i)) \)
- State-dependent observation density \( (b_j(y_t) = p(y_t | x_t = j)) \)
- Initial state probability \( (\pi_i = p(x_1 = i)) \)
Hidden Markov Models - Algorithm for Future State Probability Estimation -1

We define $\delta_t(i)$ as the normalized forward probability at time $t$ for each state $S_i$.

The state transition probabilities, $a_{ij}$ and $\delta_t(i)$ are used to predict the probabilities of states at time $t+1$. The transition probability to state $S_j$ at the time instance $t+1$ is given by:

$$[P[q_{t+1} = S_j | \lambda] = \sum_{i=1}^{j} P[q_t = S_i | \lambda]a_{ij} = \sum_{i=1}^{j} \delta_t(i)a_{ij}$$

$\lambda$ is the set of model parameters. The most probable state at time $t+1$ is the one that has the highest probability. The predicted state probabilities are updated using state dependent observation $b_t(j)$ at each time step.
Hidden Markov Models - Algorithm for Future State Probability Estimation

- Initialization: First the initial state probabilities are updated based on the freshly acquired information. $\delta$ contains the update probabilities about the machine condition at the present time. The HMM is doubly statistical, therefore the states are probabilistically estimated from the motor current.

$$\delta_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$$

$$q_1(i) = 0 \quad 1 \leq i \leq N$$

- Recursion: In this phase the future state probabilities are estimated, based on the probabilities of current state and transition probabilities. These were computed beforehand from the collected data.

$$q_t(j) = \arg \max_{1 \leq i \leq N} \sum_{1 \leq i \leq N} [\delta_{t-1}(i)a_{ij}] \quad 2 \leq t \leq T, 1 \leq j \leq N$$

Finally the state which has the largest probability at the future time is the most likely state.

$$\delta_t(j) = \sum_{1 \leq i \leq N} [\delta_{t-1}(i)a_{ij}]b_j(O_t) \quad 2 \leq t \leq T, \quad 1 \leq j \leq N$$
Hidden Markov Model – an example

• In this case determine the gear fault of a starter motor, and predict the most probable next state.

• The state dependant observation density $B$, $(b_j(O_t) = p(O_t|S_t = i))$ is the probability of observing a Feature set, given that the machine is in fault severity state $S_i$. This is

$$P(O|S_i) = \frac{P(S_i|O) \times P(O)}{P(S)}$$

These probabilities are assumed Gaussian:

$$P(O|S_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{O - \mu_{O|S_i}}{\sigma_{O|S_i}}\right)^2$$

• The statistics were obtained from the experimental data. The output of training of the LDC classifier were coefficients ($\alpha \in \mathbb{R}^{k+1}$) for each class, which gives the maximum discrimination. The sets of coefficients were used as projection hyperplanes, and each training sample, $x^j = \{x^j_1 \cdots x^j_k\}, x^j \in \mathbb{R}^k$, was projected to all $C$ planes as:

$$D_c(x^j) = x^j_1\alpha_{1c} + x^j_2\alpha_{2c} + \ldots + x^j_k\alpha_{kc} + \alpha_{k+1,c} \quad c = 1, 2, \ldots, C$$
Hidden Markov Model – an example

<table>
<thead>
<tr>
<th></th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0759</td>
<td>-0.0008</td>
<td>0.0087</td>
<td>-0.0052</td>
<td>0.0012</td>
</tr>
<tr>
<td>2</td>
<td>0.0148</td>
<td>0.0793</td>
<td>0.0213</td>
<td>0.0042</td>
<td>0.0143</td>
</tr>
<tr>
<td>3</td>
<td>0.0582</td>
<td>0.0449</td>
<td>0.1284</td>
<td>0.0533</td>
<td>0.0525</td>
</tr>
<tr>
<td>4</td>
<td>0.0665</td>
<td>0.0537</td>
<td>0.0379</td>
<td>0.2726</td>
<td>0.0223</td>
</tr>
<tr>
<td>5</td>
<td>0.2117</td>
<td>0.2161</td>
<td>0.2326</td>
<td>0.2552</td>
<td>0.4383</td>
</tr>
</tbody>
</table>

$P_i$ is the LDC plane

Means of the projection on each plane

<table>
<thead>
<tr>
<th></th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$P_4$</th>
<th>$P_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0081</td>
<td>0.0084</td>
<td>0.0103</td>
<td>0.0140</td>
<td>0.0138</td>
</tr>
<tr>
<td>2</td>
<td>0.0088</td>
<td>0.0066</td>
<td>0.0112</td>
<td>0.0101</td>
<td>0.0143</td>
</tr>
<tr>
<td>3</td>
<td>0.0231</td>
<td>0.0294</td>
<td>0.0336</td>
<td>0.0410</td>
<td>0.0358</td>
</tr>
<tr>
<td>4</td>
<td>0.1134</td>
<td>0.1282</td>
<td>0.1934</td>
<td>0.1712</td>
<td>0.2363</td>
</tr>
<tr>
<td>5</td>
<td>0.1753</td>
<td>0.1669</td>
<td>0.2522</td>
<td>0.2631</td>
<td>0.2612</td>
</tr>
</tbody>
</table>

$P_i$ is the LDC plane

Variances of the projections of the samples from each class on LDC planes
Hidden Markov Model – an example

But what about state transition probabilities, \( A \)?

Alternatives:

- Large scale testing, usually not practical
- Fatigue Analysis, complex, preferred
- Online estimation – similar to earlier transitions.

In this case, Matching Pursuit Decomposition

Transition matrix \( A \)

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5063</td>
<td>0.2435</td>
<td>0.1481</td>
<td>0.0800</td>
<td>0.0221</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.4935</td>
<td>0.2967</td>
<td>0.1751</td>
<td>0.0347</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.4542</td>
<td>0.3709</td>
<td>0.1749</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.5829</td>
<td>0.4171</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Hidden Markov Model – an example

How to test in the case of slowly evolving faults?

In this example:

- Use real data from experiments with all states,
- Create an artificial sequence of faults, and add noise to the observations,
- Observe the resulting fault progression.

![ Probable Next State vs. Observation No. ]

![ Failure State Probability vs. Observation No. ]
Fault evolution - Kalman predictor

- Select features that are the best for the application,
- Improve separation, increase compactness,
- Almost ideal situation:
  - Small within-class scatter
  - Large between-class scatter

- Limit the number of measurements
- Learn from a database a law modeling the different states

Representation of fault evolution trajectory.

Representation of membership function.
Fault evolution in rotor bars – Physics model, crack growth

Stresses next to broken bar and away from it

Paris law: \( \frac{d\alpha}{dN} = C(\Delta \sigma Y \cdot \sqrt{\pi \alpha})^m \)

Climente-Alarcón et al, "Use of high order harmonics for diagnosis of simultaneous faults via Wigner-Ville distributions," IECON 2010
Reliability

- What do all these have to do with reliability of a drive?
- What is reliability?
- Can fault Diagnosis and Prognosis improve component/subsystem/system reliability?

**Reliability:** the probability that the item will perform its required function in a stated time interval.

**Failure:** when the item stops performing its required function.

The reliability function $R(t)$ represents the probability that the item will operate without failures over a time interval $[0; t]$. 
Reliability and Failure rate

Failure rate $\lambda(t)$:

$$
\lambda(t) = \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{R(t) \Delta t} = -\frac{1}{R(t)} \frac{dR(t)}{dt}
$$

The reliability $R(t)$ then is determined from the failure rate $\lambda(t)$ with the consideration of $R(0) = 1$,

$$
R(t) = e^{-\int_0^t \lambda(\tau) d\tau}
$$

The MTTF is the expected time before a failure occurs. Unlike reliability, MTTF does not depend on a particular period of time. It gives the average time in which an item operates without failing.

$$
MTTF = \frac{1}{\lambda}
$$
Reliability of a system with many components

\[ R_s = R_1 \]

\[ R_s = R_1 R_2 R_3 \]

\[ R_s = R_1 + R_2 - R_1 R_2 \]
How does a drive fail?

- We have already identified the components that can fail.
- Operating conditions: environment and internal loads:
  - Temperature affects most components,
  - Voltage, voltage pulses, and current stresses,
  - Speed.
- Handbooks of experimentally established reliability, mostly for electronics and insulation
- Less analysis is available on mechanical faults and fatigue.
How does prognosis improve reliability

Even with very noisy and uncertain observations, prognosis improves the drive reliability. An example of intermittent opens, with the stator current $i_q$ the only measurement.

State Probability. The probability of an observation given the state.

Probability of failure state as determined directly from observations (diagnosis) and through failure prognosis.
How to increase the reliability of a drive?

- Design – overdesign
- Redundancy: sensors, inverters, motors
  - But we need to know when to use redundancy
- Internal redundancy (multi-phase machines, neutral connections,)
- Accurate decision on faults,
- Timely mitigation.
Redundancy - Types

- Active (parallel, hot): Load sharing from the beginning, equal failure rates.
- Warm redundancy: Lower load, load sharing, lower failure rate
- Standby redundancy: no load sharing, zero failure rate, time to transfer

What is the case with electrical drives?

- Of course active redundancy: e.g. a full inverter operating in parallel.
- Alternatives: A rotor position estimator, in parallel with a rotor position sensor.
- Imbedded ability to operate. N-1 instead of N phases, etc.
Challenges specific to Drives

- Typical problems of electrical machines
  - Bearings,
  - Insulation,
  - Magnets or rotor bars,
  - Eccentricity.

- Power electronics etc.
  - Capacitors
  - Switches
  - Drivers
  - Connections

- Controllers and sensors
  - Current and voltage
  - Rotor position
Which can be mitigated? What effects will this have?

- Rotor position sensor: requires controller action. Decrease in performance, increase in losses.
- One phase open: controller action, power limitation, higher temperature.
- One phase shorted: controller action, performance,
- Gears, bearings: Limited ability to compensate.
Calculate reliability of the drive
Mitigating a fault results in a drive with decreased performance and increased stresses.

Every fault determination is made with a level of certainty that has to be evaluated. Otherwise:

- A false positive:
  - Depends on the sampling rate and certainty
  - Can lead to untimely mitigation

- A false negative:
  - Will lead to inaction and either
  - Delayed mitigation and possibly secondary faults
  - Or, no mitigation at all and failure

- Appropriate thresholds would increase the reliability of a drive with mitigation
Open issues

- Despite the plethora of fault diagnosis methods, there has not been a consensus on what is appropriate at any fault type.
- Extracting state probabilities from observations, remains a challenge.
- We do not have adequate models of fault development in electrical drives (rotor bars, demagnetization, solder and welding etc.).
- Limitations of data-based models and comparisons to physics-based ones.
- We have to determine the effect of operating conditions on these models.
- Non-Markovian methods have to be developed to account for fatigue.