Short-Term Microgrid Demand Prediction Using an Ensemble of Linearly-Constrained Estimators

Md. Zulfiquar Ali Bhotto, Richard Jones, Stephen Makonin, *Senior Member, IEEE,*
and Ivan V. Bajić, *Senior Member, IEEE,*

III-*J*. Geometric Mean Combiner Node

$$\min_{\mathbf{a}_i} \ J(\mathbf{a}_i) = \mathbb{E}\left[\left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_g[i]\right)^2\right]$$

subject to the constraints

$$\sum_{l=1}^{9} \mathbf{a}_i[l] = 1, \quad \mathbf{a}_i > \mathbf{0}$$

where

$$\widehat{\mathbf{d}}_g[i] = \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_i[l]}$$

for $i = 1, 2, \ldots, 24$. For stochastic optimization we made the approximation $J(\mathbf{a}_i) \approx \hat{J}(\mathbf{a}_i) = \left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_g[i]\right)^2$ and minimize $\hat{J}(\mathbf{a}_i)$ instead of $J(\mathbf{a}_i)$. The gradient of $\hat{J}(\mathbf{a}_i)$ with respect to $\mathbf{a}_i$ becomes

$$\begin{aligned}
\mathbf{g}_i &= [\log(\widehat{\mathbf{d}}_{1,k}[i]) \ \log(\widehat{\mathbf{d}}_{2,k}[i]) \ \cdots \ \log(\widehat{\mathbf{d}}_{9,k}[i])]^\top \\
\widehat{\mathbf{d}}_g[i] &= \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_i[l]} \\
\frac{\partial \hat{J}(\mathbf{a}_i)}{\partial(\mathbf{a}_i)} &= -\mathbf{g}_i \widehat{\mathbf{d}}_g[i] \left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_g[i]\right)
\end{aligned}$$

Using the negative of the gradient we get the recursion formula

$$\begin{aligned}
\mathbf{g}_i &= [\log(\widehat{\mathbf{d}}_{1,k}[i]) \ \log(\widehat{\mathbf{d}}_{2,k}[i]) \ \cdots \ \log(\widehat{\mathbf{d}}_{9,k}[i])]^\top \\
\widehat{\mathbf{d}}_{g,k}[i] &= \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k-1}[l]} \\
\mathbf{a}_{i,k} &= \mathbf{a}_{i,k-1} + \mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] \left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right)
\end{aligned}$$

The constraint $\sum_{l=1}^{9} \mathbf{a}_i[l] = 1$ is satisfied by the using the projection matrix $\mathbf{Z}$ as

$$\begin{aligned}
\mathbf{g}_i &= [\log(\widehat{\mathbf{d}}_{1,k}[i]) \ \log(\widehat{\mathbf{d}}_{2,k}[i]) \ \cdots \ \log(\widehat{\mathbf{d}}_{9,k}[i])]^\top \\
\mathbf{Z} &= \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{9} \\
\widehat{\mathbf{d}}_{g,k}[i] &= \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k-1}[l]} \\
\mathbf{a}_{i,k} &= \mathbf{a}_{i,k-1} + \mu \mathbf{Z} \mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] \left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right)
\end{aligned}$$

Research funded by NSERC Enagage Grant EGP543219-19 in Canada.
Special thanks to Elena Popovici (elena@awesense.com) and our industry partner Awesense Wireless Inc. for their valuable feedback.

with the initialization of $\mathbf{a}_{i,0} = 1/9 \cdot \mathbf{1}$ where μ is the step size. Next we would find out the optimal step size μ. For this we premultiply both sides of the above formula by $\mathbf{g}_i^\top$ to get

$$\log\left(\prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k}[l]}\right) = \log\left(\widehat{\mathbf{d}}_{g,k}[i]\right) + \mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i]\left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right) \qquad i.e.,$$

$$\prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k}[l]} = \widehat{\mathbf{d}}_{g,k}[i] \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i]\left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right)\right)$$

Subtracting the above equation from $\mathbf{d}_k[i]$ we obtain the *a posteriori* ε_k error as

$$\varepsilon_k = \mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i] \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i]\left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right)\right)$$

We now obtain the step size μ by minimizing the cost function ε_k^2 i.e.,

$$\min_{\mu} \ \left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i] \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i]\left(\mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i]\right)\right)\right)^2$$

The derivative of the above cost function can be obtained as

$$\begin{aligned}
e_k &= \mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i] \\
\frac{\partial \varepsilon_k^2}{\partial \mu} &= -2\mathbf{d}_k[i]\widehat{\mathbf{d}}_{g,k}[i]^2 \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i e_k \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] e_k\right) \\
&\quad + 2\widehat{\mathbf{d}}_{g,k}[i]^3 \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i e_k \exp\left(2\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] e_k\right) \\
&= 2\widehat{\mathbf{d}}_{g,k}[i]^2 \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i e_k \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] e_k\right) \\
&\quad \times \left[-\mathbf{d}_k[i] + \widehat{\mathbf{d}}_{g,k}[i] \exp\left(\mu \mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i \widehat{\mathbf{d}}_{g,k}[i] e_k\right)\right]
\end{aligned}$$

Setting the derivative equal to zero we get the optimal step size to be

$$\begin{aligned}
e_k &= \mathbf{d}_k[i] - \widehat{\mathbf{d}}_{g,k}[i] \\
\mu &= \frac{1}{\widehat{\mathbf{d}}_{g,k}[i]\mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i e_k} \log\left(\frac{\mathbf{d}_k[i]}{\widehat{\mathbf{d}}_{g,k}[i]}\right)
\end{aligned}$$

Putting this step size μ back to the recursion formulas we get

$$\begin{aligned}
\mathbf{g}_i &= [\log(\widehat{\mathbf{d}}_{1,k}[i]) \ \ \log(\widehat{\mathbf{d}}_{2,k}[i]) \ \ \cdots \ \ \log(\widehat{\mathbf{d}}_{9,k}[i])]^\top \\
\mathbf{Z} &= \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{9} \\
\widehat{\mathbf{d}}_{g,k}[i] &= \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k-1}[l]} \\
\mathbf{a}_{i,k} &= \mathbf{a}_{i,k-1} + \frac{1}{\mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i} \mathbf{Z}\mathbf{g}_i \log\left(\frac{\mathbf{d}_k[i]}{\widehat{\mathbf{d}}_{g,k}[i]}\right)
\end{aligned}$$

We now have to satisfy the constraint $\mathbf{a}_i \geq 0$ which can be achieved by using

$$\mu = \frac{\min(\mathbf{a}_{i,k-1})}{\max\left(\left|\frac{1}{\mathbf{g}_i^\top \mathbf{Z}\mathbf{g}_i} \mathbf{Z}\mathbf{g}_i \log\left(\frac{\mathbf{d}_k[i]}{\widehat{\mathbf{d}}_{g,k}[i]}\right)\right|\right)}$$

in the formula of $\mathbf{a}_{i,k}$. The final set of formulas therefore becomes

$$\mathbf{g}_i = [\log(\widehat{\mathbf{d}}_{1,k}[i]) \;\; \log(\widehat{\mathbf{d}}_{2,k}[i]) \;\; \cdots \;\; \log(\widehat{\mathbf{d}}_{9,k}[i])]^\top$$

$$\mathbf{Z} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{9}$$

$$\widehat{\mathbf{d}}_{g,k}[i] = \prod_{l=1}^{9} \widehat{\mathbf{d}}_{l,k}[i]^{\mathbf{a}_{i,k-1}[l]} \tag{22}$$

$$e_i = \log\left(\frac{\mathbf{d}_k[i]}{\widehat{\mathbf{d}}_{g,k}[i]}\right)$$

$$\mathbf{q}_i = \mathbf{Z}\mathbf{g}_i$$

$$\mu = \frac{\min(\mathbf{a}_{i,k-1})}{\max(|\mathbf{q}_i e_i|)}$$

$$\mathbf{a}_{i,k} = \mathbf{a}_{i,k-1} + \mu \mathbf{q}_i e_i \tag{23}$$

for $i = 1, 2, \ldots, 24$ and $k = 1, 2, \ldots, N$.

III-*I*. Alternating Minimization Based Linear Prediction Node

The cost function in this section is

$$\min_{\mathbf{w}_9, \mathcal{W}_2} \; J(\mathbf{w}_9, \mathcal{W}_2) = \mathbb{E}\left[\|\mathbf{d}_k - \mathcal{W}_2 \mathbf{D}_k \mathbf{w}_9\|^2\right]$$

For stochastic optimization we made the approximation $J(\mathbf{w}_9, \mathcal{W}_2) \approx \hat{J}(\mathbf{w}_9, \mathcal{W}_2) = \|\mathbf{d}_k - \mathcal{W}_2 \mathbf{D}_k \mathbf{w}_9\|^2$ and minimize $\hat{J}(\mathbf{w}_9, \mathcal{W}_2)$ instead of $J(\mathbf{w}_9, \mathcal{W}_2)$. The gradients of $\hat{J}(\mathbf{w}_9, \mathcal{W}_2)$ with respect to $\mathbf{w}_9$ and $\mathcal{W}_2$ become

$$\mathbf{y} = \mathbf{D}_k \mathbf{w}_9$$

$$\frac{\partial \hat{J}(\mathbf{w}_9, \mathcal{W}_2)}{\partial \mathcal{W}_2} = -\left(\mathbf{d}_k - \mathcal{W}_2 \mathbf{y}\right) \mathbf{y}^\top$$

$$\frac{\partial \hat{J}(\mathbf{w}_9, \mathcal{W}_2)}{\partial \mathbf{w}_9} = -\mathbf{D}_k^\top \mathcal{W}_2^\top \left(\mathbf{d}_k - \mathcal{W}_2 \mathbf{y}\right)$$

By using the negative of these gradients we obtain their update formulas as

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{D}_k \mathbf{w}_{9,k-1} \\
\widehat{\mathbf{d}}_{9,k} &= \mathcal{W}_{2,k-1} \mathbf{y}_k \\
\mathbf{e}_k &= \mathbf{d}_k - \widehat{\mathbf{d}}_{9,k} \\
\mathbf{q}_{1,k} &= \mathbf{D}_k^\top \mathcal{W}_{2,k-1}^\top \mathbf{e}_k \\
\mathbf{q}_{2,k} &= \mathbf{e}_k \mathbf{y}_k^\top \\
\mathbf{w}_{9,k} &= \mathbf{w}_{9,k-1} + \mu_1 \mathbf{q}_{1,k} \\
\mathcal{W}_{2,k} &= \mathcal{W}_{2,k-1} + \mu_2 \mathbf{q}_{2,k}
\end{aligned} \tag{21}$$

where μ_1 and μ_2 are the two step sizes.

III-*H*. Kurtosis Based Prediction Node

The cost function in this node is

$$\min_{\mathbf{w}_8} \; \max_{\mathcal{W}_1} \; J(\mathbf{w}_8, \mathcal{W}_1) = \mathbb{E}\left[\mathbf{e}^\top (\mathcal{W}_1 - \mathbf{e}\mathbf{e}^\top) \mathbf{e}\right]$$

where $\mathbf{e} = \mathbf{d}_k - \mathbf{D}_k \mathbf{w}_8$. For stochastic optimization we made the approximation $J(\mathbf{w}_8, \mathcal{W}_1) \approx \hat{J}(\mathbf{w}_8, \mathcal{W}_1) = \mathbf{e}^\top (\mathcal{W}_1 - \mathbf{e}\mathbf{e}^\top)\mathbf{e}$. The gradients of $\hat{J}(\mathbf{w}_8, \mathcal{W}_1)$ with respect to $\mathbf{w}_8$ and $\mathcal{W}_1$ become

$$\mathbf{e} = \mathbf{d}_k - \mathbf{D}_k \mathbf{w}_8$$

$$\frac{\partial \hat{J}(\mathbf{w}_8, \mathcal{W}_1)}{\partial \mathcal{W}_1} = \mathbf{e}\mathbf{e}^\top$$

$$\frac{\partial \hat{J}(\mathbf{w}_8, \mathcal{W}_1)}{\partial \mathbf{w}_8} = -\mathbf{D}_k^\top (\mathcal{W}_1 - 2\mathbf{e}\mathbf{e}^\top)\mathbf{e}$$

By using the gradient $\partial\mathcal{W}_1$ as it is and the negative of the gradient $\partial\mathbf{w}_8$ we obtain their update formulas as

$$\begin{aligned}\widehat{\mathbf{d}}_{8,k} &= \mathbf{D}_k\mathbf{w}_{8,k-1}\\ \mathbf{e}_k &= \mathbf{d}_k - \widehat{\mathbf{d}}_{8,k}\\ \mathcal{W}_e &= 3\mathcal{W}_{1,k-1} - \mathbf{e}_k\mathbf{e}_k^\top\\ \mathcal{W}_{1,k} &= \lambda\mathcal{W}_{1,k-1} + \beta\mathbf{e}_k\mathbf{e}_k^\top\\ \mathbf{w}_{8,k} &= \mathbf{w}_{8,k-1} + \mu\mathbf{D}_k^\top\mathcal{W}_e\mathbf{e}_k\end{aligned} \tag{19}$$

We obtain the step size μ by solving the cost function

$$\min_{\mu}\ \|\mathbf{d}_k - \mathbf{D}_k\mathbf{w}_{8,k}\|^2$$

as

$$\mu = \frac{\mathbf{e}_k^\top\mathbf{D}_k\mathbf{D}_k^\top\mathcal{W}_e\mathbf{e}_k}{\mathbf{e}_k^\top\mathcal{W}_e^\top\mathbf{D}_k\mathbf{D}_k^\top\mathbf{D}_k\mathbf{D}_k^\top\mathcal{W}_e\mathbf{e}_k}$$

III-*H*. MinMax Linear Prediction Node

The cost function in this node is

$$\min_{\mathbf{w}_7}\ \max_{\mathbf{c}}\ J(\mathbf{w}_7,\mathbf{c}) = \mathbb{E}\left[(\mathbf{c}^\top\mathbf{d}_k - \mathbf{c}^\top\mathbf{D}_k\mathbf{w}_7)^2\right]$$

subject to

$$\mathbf{c}\geq\mathbf{0}\ \ \text{and}\ \ \mathbf{1}^\top\mathbf{c} = 1.$$

For stochastic optimization we made the approximation $J(\mathbf{w}_7,\mathbf{c}) \approx \hat{J}(\mathbf{w}_7,\mathbf{c}) = (\mathbf{c}^\top\mathbf{d}_k - \mathbf{c}^\top\mathbf{D}_k\mathbf{w}_7)^2$. The gradients of $\hat{J}(\mathbf{w}_7,\mathbf{c})$ with respect to $\mathbf{w}_7$ and $\mathbf{c}$ become

$$\begin{aligned}\mathbf{e} &= \mathbf{d}_k - \mathbf{D}_k\mathbf{w}_7\\ \frac{\partial\hat{J}(\mathbf{w}_7,\mathbf{c})}{\partial\mathbf{c}} &= \mathbf{e}\mathbf{c}^\top\mathbf{e}\\ \frac{\partial\hat{J}(\mathbf{w}_7,\mathbf{c})}{\partial\mathbf{w}_7} &= -\mathbf{D}_k^\top\mathbf{c}\mathbf{c}^\top\mathbf{e}\end{aligned}$$

Using gradient $\partial\mathbf{c}$ and negative of the gradient $\partial\mathbf{w}_7$ we obtain the update formulas as

$$\begin{aligned}\widehat{\mathbf{d}}_k &= \mathbf{D}_k\mathbf{w}_{7,k-1}\\ \mathbf{e}_k &= \mathbf{d}_k - \widehat{\mathbf{d}}_k\\ \varepsilon_a &= \mathbf{c}_{k-1}^\top\mathbf{e}_k\\ \mathbf{c}_k &= \mathbf{c}_{k-1} + \mu_1\mathbf{e}_k\varepsilon_a\\ \mathbf{w}_{7,k} &= \mathbf{w}_{7,k-1} + \mu_2\mathbf{D}_k^\top\mathbf{c}_{k-1}\varepsilon_a\end{aligned}$$

Now to satisfy the constraint $\mathbf{1}^\top\mathbf{c} = 1$ we premultiply the second term on the right hand side of the update formula $\mathbf{c}_k$ by matrix

$$\mathbf{Z} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{24}$$

and to satisfy the constraint $\mathbf{c}\geq\mathbf{0}$ we use

$$\mu_1 = \frac{\min(\mathbf{c}_{k-1})}{\max(|\mathbf{Z}\mathbf{e}_k\varepsilon_a|)}$$

On the other hand we obtain μ_2 by solving the cost function

$$\min_{\mu_2}\ [\mathbf{c}_{k-1}^\top(\mathbf{d}_k - \mathbf{D}_k\mathbf{w}_{7,k})]^2$$

as

$$\mu_2 = \frac{1}{\|\mathbf{D}_k^\top\mathbf{c}_{k-1}\|^2}$$

The update formulas therefore assume the form

$$
\begin{aligned}
\widehat{\mathbf{d}}_k &= \mathbf{D}_k \mathbf{w}_{7,k-1} \\
\mathbf{e}_k &= \mathbf{d}_k - \widehat{\mathbf{d}}_k \\
\varepsilon_a &= \mathbf{c}_{k-1}^\top \mathbf{e}_k \\
\mathbf{Z} &= \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{24} \\
\mathbf{q}_k &= \mathbf{Z}\mathbf{e}_k \varepsilon_a \\
\mu &= \frac{\min(\mathbf{c}_{k-1})}{\max(|\mathbf{q}_k|)} \\
\mathbf{c}_k &= \mathbf{c}_{k-1} + \mu \mathbf{q}_k \\
\mathbf{w}_{7,k} &= \mathbf{w}_{7,k-1} + \frac{1}{\mathbf{c}_{k-1}^\top \mathbf{D}_k \mathbf{D}_k^\top \mathbf{c}_{k-1}} \mathbf{D}_k^\top \mathbf{c}_{k-1} \varepsilon_a
\end{aligned} \tag{17}
$$

III-*C*. Hourly Load Variation Constrained Prediction Node

The cost function in this node was

$$\min_{\mathbf{w}_3} \; J(\mathbf{w}_3) \; \mathbb{E}\left[\|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_3\|^2 + \gamma \|\mathbf{F}\mathbf{D}_k \mathbf{w}_3\|_1\right] \tag{6}$$

For stochastic optimization we made the approximation $J(\mathbf{w}_3) \approx \hat{J}(\mathbf{w}_3) = \|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_3\|^2 + \gamma \|\mathbf{F}\mathbf{D}_k \mathbf{w}_3\|_1$. We break down the cost function $\hat{J}(\mathbf{w}_3)$ in two parts as

$$\min_{\mathbf{w}_3} \; \hat{J}_1(\mathbf{w}_3) = \|\mathbf{d}_k - \mathbf{F}^{-1}\mathbf{z}_3^*\|^2$$

where

$$\mathbf{z}_3^* := \min_{\mathbf{z}_3} \; \hat{J}_2(\mathbf{z}_3) = \|\mathbf{F}\mathbf{D}_k \mathbf{w}_3 - \mathbf{z}_3\|^2 + \gamma \|\mathbf{z}_3\|_1$$

The solution of the second cost function $\hat{J}_2(\mathbf{z}_3)$ becomes

$$\mathbf{z}_3^* = \max(|\mathbf{F}\mathbf{D}_k \mathbf{w}_3| - \gamma, 0) \odot \operatorname{sign}(\mathbf{F}\mathbf{D}_k \mathbf{w}_3)$$

The gradient of the first cost function $\hat{J}_1(\mathbf{w}_3)$ becomes

$$
\begin{aligned}
\frac{\partial \hat{J}_1(\mathbf{w}_3)}{\partial \mathrm{w}_3} &= -\frac{\partial \mathbf{F}^{-1}\mathbf{z}_3^*}{\partial \mathrm{w}_3}\left(\mathbf{d}_k - \mathbf{F}^{-1}\mathbf{z}_3^*\right) \\
&= -\mathbf{D}_k^\top \left(\mathbf{d}_k - \mathbf{F}^{-1}\mathbf{z}_3^*\right)
\end{aligned}
$$

where we avoid the cases of $\mathbf{z}_3^* = 0$ by using a small enough γ on the order of a fraction of $\mathbb{E}[\|\mathbf{F}\mathbf{d}_k\|_1]$. Using the negative of the gradient we obtain update formula

$$
\begin{aligned}
\widehat{\mathbf{d}}_{3,k} &= \mathbf{D}_k \mathbf{w}_{3,k-1} \\
\mathbf{f}_d &= \mathbf{F}\widehat{\mathbf{d}}_{3,k} \\
\mathbf{y}_k &= \mathbf{F}^{-1} \left[\mathbf{diag}(\operatorname{abs}(\mathbf{f}_d)) - \gamma\right]_+ \operatorname{sign}(\mathbf{f}_d) \\
\mathbf{e}_z &= \mathbf{d}_{3,k} - \mathbf{y}_k \\
\mathbf{q}_k &= \mathbf{D}_k^\top \mathbf{e}_z \\
\mathbf{w}_{3,k} &= \mathbf{w}_{3,k-1} + \mu \mathbf{q}_k
\end{aligned}
$$

The step size is obtained by minimizing the cost function

$$\min_{\mu} \; \|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_{3,k}\|^2$$

as

$$
\begin{aligned}
\mathbf{e}_k &= \mathbf{d}_{3,k} - \widehat{\mathbf{d}}_{3,k} \\
\mu &= \frac{\mathbf{e}_k^\top \mathbf{D}_k \mathbf{q}_k}{\|\mathbf{D}_k \mathbf{q}_k\|^2}
\end{aligned}
$$

The final set of recursions therefore become

$$\begin{aligned}
\widehat{\mathbf{d}}_{3,k} &= \mathbf{D}_k \mathbf{w}_{3,k-1} &(7)\\
\mathbf{e}_k &= \mathbf{d}_{3,k} - \widehat{\mathbf{d}}_{3,k}\\
\mathbf{f}_d &= \mathbf{F}\widehat{\mathbf{d}}_{3,k}\\
\mathbf{y}_k &= \mathbf{F}^{-1}\left[\mathbf{diag}(\mathrm{abs}(\mathbf{f}_d)) - \gamma\right]_+ \mathrm{sign}(\mathbf{f}_d)\\
\mathbf{e}_z &= \mathbf{d}_{3,k} - \mathbf{y}_k\\
\mathbf{q}_k &= \mathbf{D}_k^\top \mathbf{e}_z\\
\mu &= \frac{\mathbf{e}_k^\top \mathbf{D}_k \mathbf{q}_k}{\|\mathbf{D}_k \mathbf{q}_k\|^2}\\
\mathbf{w}_{3,k} &= \mathbf{w}_{3,k-1} + \mu \mathbf{q}_k &(8)
\end{aligned}$$

The derivation of DLP algorithm in (11) is the same as the UP algorithm in (2).

ROBUST LINEAR PREDICTION NODE

The cost function in this node was

$$\min_{\mathbf{w}_5} \ \mathbb{E}\left[\|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_5\|_1\right] \tag{1}$$

For stochastic optimization we made the approximation $J(\mathbf{w}_5) \approx \hat{J}(\mathbf{w}_5) = \|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_5\|_1$. The gradient of the cost function becomes

$$\frac{\partial \hat{J}(\mathbf{w}_5)}{\partial \mathbf{w}_5} = -\mathbf{D}^\top \mathrm{sign}(\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_5)$$

The update formula can be obtained by taking the negative of the gradient as

$$\mathbf{w}_{5,k} = \mathbf{w}_{5,k-1} + \mu \mathbf{D}^\top \mathrm{sign}(\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_{5,k-1}) \tag{2}$$

We now obtain the step size μ by solving the cost function

$$\min_{\mu} \ \|\mathbf{d}_k - \mathbf{D}_k \mathbf{w}_{5,k}\|^2$$

as

$$\mu = \alpha \frac{\mathbf{e}_k^\top \mathbf{D}_k \mathbf{D}_k^\top \mathrm{sign}(\mathbf{e}_k)}{\mathrm{sign}(\mathbf{e}_k)^\top \mathbf{D}_k \mathbf{D}_k^\top \mathbf{D}_k \mathbf{D}_k^\top \mathrm{sign}(\mathbf{e}_k)}$$

where we use $0 < \alpha \ll 1$ to reduce the amplitude of the absolute error. The RdP algorithm is straightforward extension of RbP and hence not presented. This completes all the derivations of the algorithms used in different nodes.