Supplementary material for **A Unified View of Multi-Label Performance Measures**.

A Appendix A: Proofs

In this section, we provide detailed proofs of the theoretical results presented in the manuscript.

A.1 Proof of Theorem 4

Because F is label-wise effective, its order of prediction value on a specific instance x_i is correct. Therefore, the threshold error ϵ_i can happen in either the two ways:

1. ϵ_i positive labels are predicted as negative labels.
   In this case, the true positive number TP_i on this instance becomes $|Y_{i\cdot}^+| - \epsilon_i$, and the false positive number FP_i is zero, and the false negative number FN_i becomes ϵ_i.
   The precision value and the recall value will be:
   $$Prec_i = \frac{TP_i}{TP_i + FP_i} = 1, \quad Rec_i = \frac{TP_i}{TP_i + FN_i} = \frac{|Y_{i\cdot}^+| - \epsilon_i}{|Y_{i\cdot}^+|}$$
   And the F-measure$_i$ is:
   $$F\text{-measure}_i = \frac{2Prec_i \times Rec_i}{Prec_i + Rec_i} = \frac{2(|Y_{i\cdot}^+| - \epsilon_i)}{2|Y_{i\cdot}^+| - \epsilon_i}$$
2. ϵ_i negative labels are predicted as positive labels.
   In this case, the true positive number TP_i on this instance is still $|Y_{i\cdot}^+|$, and the false positive number $FP_i = \epsilon_i$, and the false negative number FN_i is zero.
   The precision value and the recall value will be:
   $$Prec_i = \frac{TP_i}{TP_i + FP_i} = \frac{|Y_{i\cdot}^+|}{|Y_{i\cdot}^+| + \epsilon_i}, \quad Rec_i = \frac{TP_i}{TP_i + FN_i} = 1$$
   And the F-measure$_i$ is:
   $$F\text{-measure}_i = \frac{2Prec_i \times Rec_i}{Prec_i + Rec_i} = \frac{2|Y_{i\cdot}^+|}{2|Y_{i\cdot}^+| + \epsilon_i}$$

The instance-F1 is lower bounded by the sum of minimum value of F-measure$_i$, thus:

$$\textit{instance-F1}(H) \geq \frac{1}{m}\sum_{i=1}^{m} \min\left\{\frac{2(|Y_{i\cdot}^+| - \epsilon_i)}{2|Y_{i\cdot}^+| - \epsilon_i}, \frac{2|Y_{i\cdot}^+|}{2|Y_{i\cdot}^+| + \epsilon_i}\right\}$$

Under the assumption that all the instances are i.i.d drawn, micro-F1 equals instance-F1. Theorem 4 is proved. □

A.2 Proof of Theorem 5

Because F is instance-wise effective, its order of prediction value on a specific label $\boldsymbol{Y}_{\cdot j}$ is correct. Therefore, the threshold error ϵ_j can happen in either the two ways:

1. ϵ_j positive instances are predicted as negative instances.
   In this case, the true positive number TP_j on this label becomes $|Y_{\cdot j}^+| - \epsilon_j$, and the false positive number FP_j is zero, and the false negative number FN_j becomes ϵ_j.
   The precision value and the recall value will be:
   $$Prec_j = \frac{TP_j}{TP_j + FP_j} = 1, \quad Rec_j = \frac{TP_j}{TP_j + FN_j} = \frac{|Y_{\cdot j}^+| - \epsilon_j}{|Y_{\cdot j}^+|}$$
   And the F-measure$_j$ is:
   $$F\text{-measure}_j = \frac{2Prec_j \times Rec_j}{Prec_j + Rec_j} = \frac{2(|Y_{j\cdot}^+| - \epsilon_j)}{2|Y_{j\cdot}^+| - \epsilon_j}$$

2. ϵ_j negative instanaces are predicted as positive instances.
In this case, the true positive number TP_j on this label is still $|Y_{\cdot j}^+|$, and the false positive number $FP_j = \epsilon_j$, and the false negative number FN_j is zero.
The precision value and the recall value will be:

$$Prec_j = \frac{TP_j}{TP_j + FP_j} = \frac{|Y_{\cdot j}^+|}{|Y_{\cdot j}^+| + \epsilon_j}, \quad Rec_j = \frac{TP_j}{TP_j + FN_j} = 1$$

And the F-measure$_j$ is:

$$F\text{-measure}_j = \frac{2Prec_j \times Rec_j}{Prec_j + Rec_j} = \frac{2|Y_{\cdot j}^+|}{2|Y_{\cdot j}^+| + \epsilon_j}$$

The macro-F1 is lower bounded by the sum of minimum value of F-measure$_j$, thus:

$$\textit{macro-F1}(H) \geq \frac{1}{l}\sum_{j=1}^{l} \min\left\{\frac{2(|Y_{\cdot j}^+| - \epsilon_j)}{2|Y_{\cdot j}^+| - \epsilon_j}, \frac{2|Y_{\cdot j}^+|}{2|Y_{\cdot j}^+| + \epsilon_j}\right\}$$

Theorem 5 is proved. □

A.3 Proof of LIMO Algorithm

Theorem A.3.1. *In each iteration (step 5 to step 15) of Algorithm 1, the updated direction of the model is an unbiased estimation of the gradient of this objective function:*

$$\begin{aligned}
&\arg\min_{\boldsymbol{W},\xi} \sum_{i=1}^{l} ||\boldsymbol{w}_i||^2 + \lambda_1 \sum_{i=1}^{m}\sum_{(u,v)} \xi_i^{uv} + \lambda_2 \sum_{j=1}^{l}\sum_{(a,b)} \xi_{ab}^{j} \\
&\quad s.t.\ \boldsymbol{w}_u^\top \boldsymbol{x}_i - \boldsymbol{w}_v^\top \boldsymbol{x}_i > 1 - \xi_i^{uv},\ \ \xi_i^{uv} \geq 0,\ for\ i = 1, \cdots, m\ and\ (u,v) \in Y_{i\cdot}^+ \times Y_{i\cdot}^-\ , \\
&\quad \boldsymbol{w}_j^\top \boldsymbol{x}_a - \boldsymbol{w}_j^\top \boldsymbol{x}_b > 1 - \xi_{ab}^{j},\ \ \xi_{ab}^{j} \geq 0,\ for\ j = 1, \cdots, l\ and\ (a,b) \in Y_{\cdot j}^+ \times Y_{\cdot j}^-\ .
\end{aligned} \tag{1}$$

Proof. Suppose that the function in Equation (1) is $f(\boldsymbol{W})$, because $\boldsymbol{W}$ can be decomposed into $[\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_l]$, we consider the partial gradient of a particular $\boldsymbol{w}_k$:

$$\begin{aligned}
\frac{\partial f(\boldsymbol{W})}{\partial \boldsymbol{w}_k} =& 2\boldsymbol{w}_k + \lambda_1\phi_1 + \lambda_2\phi_2 = 2\boldsymbol{w}_k \\
&+ \lambda_1 \sum_{i=1}^{m} \Big\{ [\![k \in Y_{i\cdot}^-]\!] \boldsymbol{x}_i \sum_{j \in Y_{i\cdot}^+} [\![1 - (\boldsymbol{w}_j - \boldsymbol{w}_k)^\top \boldsymbol{x}_i > 0]\!] \\
&\qquad - [\![k \in Y_{i\cdot}^+]\!] \boldsymbol{x}_i \sum_{j \in Y_{i\cdot}^-} [\![1 - (\boldsymbol{w}_k - \boldsymbol{w}_j)^\top \boldsymbol{x}_i > 0]\!] \Big\} \\
&+ \lambda_2 \sum_{a \in Y_{\cdot k}^+} \sum_{b \in Y_{\cdot k}^-} (\boldsymbol{x}_b - \boldsymbol{x}_a) [\![1 - \boldsymbol{w}_k^\top (\boldsymbol{x}_a - \boldsymbol{x}_b) > 0]\!]
\end{aligned} \tag{2}$$

The second term $\lambda_1\phi_1$ is the gradient of label-wise margin on $\boldsymbol{w}_k$, and the third term $\lambda_2\phi_2$ is the gradient of the instance-wise margin on $\boldsymbol{w}_k$.

Assume that $(\boldsymbol{x}_i, y_{ik}, y_{ij})$ is picked in step 5 and 6, the direction will be computed in step 8 or 9 according to:

$$\begin{aligned}
g^{label}(\boldsymbol{x}_i, y_{ik}, y_{ij}) =& [\![k \in Y_{i\cdot}^-]\!] \lambda_1 \boldsymbol{x}_i [\![1 - (w_k - w_j)^\top \boldsymbol{x}_i > 0]\!] \\
&- [\![k \in Y_{i\cdot}^+]\!] \lambda_1 \boldsymbol{x}_i [\![1 - (w_j - w_k)^\top \boldsymbol{x}_i > 0]\!] + \boldsymbol{w}_k
\end{aligned}$$

Then do the expectation:

$$\begin{aligned} E_{\boldsymbol{x}_i}\left[E_{y_{ij}}[g^{label}(\boldsymbol{x}_i, y_{ik}, y_{ij})]\right] &= \frac{1}{C} E_{\boldsymbol{x}_i}\Bigg[\lambda_1 \boldsymbol{x}_i \sum_{j \in Y_{i\cdot}^+} [\![1 - (w_k - w_j)^\top \boldsymbol{x}_i > 0]\!] \\ &\qquad - \lambda_1 \boldsymbol{x}_i \sum_{j \in Y_{i\cdot}^-} [\![1 - (w_j - w_k)^\top \boldsymbol{x}_i > 0]\!] + \frac{1}{D} \boldsymbol{w}_k \Bigg] \\ &= \frac{1}{C'} \lambda_1 \phi_1 + \frac{1}{D'} \boldsymbol{w}_k \end{aligned}$$

Where C' and D' are constants. Similarly, we can prove the expectation of the direction in step 11 to 15:

$$E_{\boldsymbol{x}_a, \boldsymbol{x}_b}[g^{inst}(y_k, \boldsymbol{x}_a, \boldsymbol{x}_b)] = \frac{1}{C''} \lambda_2 \phi_2 + \frac{1}{D''} \boldsymbol{w}_k$$

Because of the linearity of expectation, and absorbing the constants into λ_1 and λ_2, the gradient $\frac{\partial f(\boldsymbol{W})}{\partial \boldsymbol{w}_k}$ can be unbiased estimated. Namely, the updated direction of the algorithm is an unbiased estimation of the gradient of Equation (1). $\square$

B Appendix B: Detailed Experimental Results

In this section, detailed experimental results are included. The results of synthetic data are in Section B.1 and The results of benchmark data are in Section B.2

B.1 Detailed Experimental Results of Synthetic Data

In this section, the detailed experimental results of synthetic data are given.

Table B.1: Original absolute value and rescaled value of experiments on ranking measures. In the left columns are absolute values, and in the right columns are rescaled relative values.

measure	LIMO-inst		LIMO		LIMO	
ranking loss	0.027	0.00	0.015	0.99	0.015	1.00
avg. precision	0.992	0.00	0.992	0.58	0.992	1.00
one-error	0.000	1.00	0.001	0.28	0.001	0.00
coverage	1.576	0.00	1.557	0.97	1.556	1.00
macro-AUC	0.842	1.00	0.828	0.00	0.842	0.98
instance-AUC	0.973	0.00	0.985	0.99	0.985	1.00
micro-AUC	0.861	0.14	0.854	0.00	0.903	1.00

Table B.2: Original absolute value and rescaled value of experiments on classification measures. In the left columns are absolute values, and in the right columns are rescaled relative values.

measure	LIMO-inst-t		LIMO-inst-t(x)		LIMO-label-t		LIMO-label-t(x)		LIMO-t		LIMO-t(x)	
Hamming loss	0.172	0.28	0.160	0.48	0.188	0.00	0.131	1.00	0.163	0.43	0.134	0.94
micro-F1	0.837	0.00	0.860	0.43	0.840	0.06	0.890	1.00	0.858	0.40	0.885	0.92
macro-F1	0.869	0.87	0.857	0.25	0.861	0.46	0.859	0.35	0.872	1.00	0.852	0.00
instance-F1	0.804	0.00	0.883	0.79	0.835	0.32	0.904	1.00	0.858	0.54	0.900	0.96

B.2 Detailed Experimental Results of Benchmark Data

The ranking results in Figure 4 in paper are computed from Table B.3. Because this table is too large, we can only rotate it to show in the next page.

Table B.3: Experimental results on eleven multi-label performance measures. For each performance measure, "↓" indicates "the smaller the better" and "↑" indicates "the larger the better". The results are shown in mean±std(rank). The smaller the rank, the better the performance.

Dataset	Algorithm	hamming loss↓	ranking loss↓	avg. precision↑	one-error↓	coverage↓	instance-F1↑	instance-AUC↑	macro-F1↑	macro-AUC↑	micro-F1↑	micro-AUC ↑
CAL500	BR	.145±.003(5)	.216±.005(4)	.470±.008(4)	.212±.025(5)	143.025±2.319(4)	.354±.009(5)	.784±.005(4)	.097±.006(5)	.544±.012(2)	.357±.011(5)	.779±.006(4)
	ML-kNN	.139±.003(3)	.184±.005(3)	.491±.007(3)	.106±.023(3)	129.789±2.426(2)	.321±.010(6)	.816±.005(3)	.053±.002(6)	.523±.009(3)	.318±.010(6)	.813±.004(3)
	GFM	.200±.002(6)	.522±.007(5)	.337±.004(5)	.000±.000(1)	166.481±0.906(6)	.454±.006(3)	.662±.006(5)	.183±.005(3)	.518±.013(5)	.457±.006(3)	.661±.006(5)
	LIMO-inst	.143±.004(4)	.545±.015(6)	.147±.004(6)	.971±.022(6)	162.652±1.539(5)	.386±.010(4)	.455±.015(6)	.302±.008(1)	.566±.011(1)	.389±.010(4)	.458±.014(6)
	LIMO-label	.138±.002(2)	.180±.004(2)	.499±.008(2)	.105±.023(2)	129.993±2.491(3)	.473±.004(2)	.820±.004(2)	.126±.003(4)	.510±.011(6)	.477±.004(2)	.815±.004(2)
	LIMO	.137±.025(1)	.178±.004(1)	.501±.008(1)	.122±.035(4)	129.323±2.672(1)	.475±.006(1)	.822±.004(1)	.288±.006(2)	.523±.011(4)	.479±.006(1)	.816±.004(1)
medical	BR	.011±.001(1)	.073±.041(5)	.416±.100(6)	.804±.029(6)	3.697±1.752(5)	.766±.022(1)	.927±.040(5)	.384±.040(3)	.877±.038(3)	.792±.020(1)	.910±.039(5)
	ML-kNN	.016±.001(5)	.048±.008(4)	.788±.017(4)	.266±.025(5)	3.034±0.411(4)	.564±.033(5)	.953±.007(4)	.190±.015(6)	.797±.029(5)	.654±.028(4)	.949±.008(4)
	GFM	.025±.002(6)	.287±.026(6)	.692±.025(5)	.217±.025(3)	6.581±0.714(6)	.636±.025(4)	.882±.013(6)	.216±.018(4)	.650±.027(6)	.605±.027(5)	.875±.014(6)
	LIMO-inst	.015±.001(4)	.017±.005(1)	.881±.018(2)	.170±.028(2)	1.248±0.279(1)	.444±.090(6)	.983±.005(1)	.448±.024(2)	.901±.032(1)	.439±.087(6)	.979±.005(1)
	LIMO-label	.014±.001(3)	.032±.006(3)	.829±.018(3)	.217±.026(4)	2.237±0.378(3)	.641±.030(3)	.968±.006(3)	.207±.012(5)	.859±.035(4)	.702±.020(3)	.960±.007(3)
	LIMO	.013±.001(2)	.019±.006(2)	.893±.017(1)	.147±.027(1)	1.423±0.350(2)	.706±.019(2)	.981±.006(2)	.464±.024(1)	.896±.029(2)	.757±.012(2)	.977±.006(2)
enron	BR	.070±.003(6)	.136±.010(4)	.539±.086(4)	.533±.263(6)	16.834±0.669(4)	.482±.009(3)	.866±.010(4)	.187±.015(3)	.631±.025(5)	.473±.008(3)	.814±.008(4)
	ML-kNN	.053±.001(3)	.096±.004(3)	.624±.014(3)	.310±.022(4)	13.615±0.423(3)	.409±.022(5)	.904±.004(3)	.083±.008(6)	.633±.022(4)	.461±.017(4)	.898±.003(1)
	GFM	.069±.003(5)	.554±.027(6)	.399±.021(6)	.246±.041(2)	31.645±0.764(6)	.428±.022(4)	.669±.014(6)	.118±.012(5)	.553±.015(6)	.437±.016(5)	.654±.011(6)
	LIMO-inst	.054±.001(4)	.205±.008(5)	.520±.008(5)	.344±.013(5)	23.679±0.804(5)	.404±.053(6)	.796±.008(5)	.310±.018(1)	.717±.015(1)	.414±.056(6)	.810±.006(5)
	LIMO-label	.049±.001(2)	.085±.003(2)	.672±.010(2)	.233±.017(1)	12.324±0.444(2)	.565±.011(1)	.916±.003(2)	.137±.005(4)	.644±.019(3)	.591±.008(2)	.897±.003(2)
	LIMO	.049±.001(1)	.083±.003(1)	.672±.010(1)	.253±.022(3)	11.880±0.255(1)	.562±.010(2)	.918±.003(1)	.278±.017(2)	.663±.021(2)	.596±.006(1)	.896±.004(3)
corel5k	BR	.014±.000(5)	.280±.010(5)	.077±.013(6)	.962±.004(6)	207.643±3.477(5)	.139±.005(4)	.720±.010(5)	.044±.003(4)	.605±.004(4)	.158±.006(3)	.706±.009(5)
	ML-kNN	.009±.000(1)	.135±.002(3)	.245±.004(2)	.736±.009(3)	114.727±1.658(3)	.017±.002(6)	.865±.002(3)	.009±.001(6)	.540±.007(5)	.027±.003(6)	.866±.002(3)
	GFM	.021±.001(6)	.803±.012(6)	.100±.005(5)	.516±.030(1)	320.449±2.173(6)	.150±.008(3)	.416±.012(6)	.029±.002(5)	.516±.006(6)	.146±.011(4)	.411±.013(6)
	LIMO-inst	.010±.000(2)	.275±.004(4)	.105±.004(4)	.897±.006(5)	172.120±2.183(4)	.058±.003(5)	.725±.004(4)	.118±.003(1)	.706±.006(1)	.057±.002(5)	.725±.004(4)
	LIMO-label	.011±.000(4)	.112±.003(1)	.289±.006(1)	.710±.010(2)	99.629±2.136(2)	.214±.004(1)	.888±.003(1)	.050±.002(3)	.658±.006(3)	.236±.004(1)	.882±.002(1)
	LIMO	.011±.000(3)	.116±.003(2)	.227±.005(3)	.791±.008(4)	94.253±2.109(1)	.152±.005(2)	.884±.003(2)	.117±.004(2)	.692±.007(2)	.177±.005(2)	.881±.003(2)
bibtex	BR	.016±.001(5)	.114±.007(3)	.528±.012(2)	.428±.014(2)	32.758±1.929(4)	.399±.009(1)	.886±.007(3)	.318±.016(3)	.866±.007(4)	.419±.012(3)	.869±.007(4)
	ML-kNN	.014±.000(2)	.218±.004(5)	.339±.006(5)	.599±.007(6)	56.259±1.260(5)	.160±.007(6)	.782±.004(5)	.066±.006(6)	.661±.007(5)	.211±.007(5)	.776±.005(5)
	GFM	.037±.000(6)	.707±.003(6)	.210±.004(6)	.492±.010(5)	85.281±0.701(6)	.223±.008(5)	.618±.003(6)	.130±.007(5)	.575±.003(6)	.185±.007(6)	.626±.003(6)
	LIMO-inst	.014±.000(1)	.120±.003(4)	.494±.008(4)	.469±.016(4)	32.403±0.640(3)	.392±.005(2)	.880±.003(4)	.323±.005(2)	.921±.002(2)	.438±.006(1)	.877±.003(3)
	LIMO-label	.014±.000(4)	.071±.002(2)	.527±.008(3)	.433±.013(3)	20.425±0.422(2)	.386±.007(4)	.929±.002(2)	.232±.004(4)	.911±.002(3)	.405±.007(4)	.917±.002(2)
	LIMO	.014±.000(3)	.058±.001(1)	.570±.004(1)	.390±.008(1)	17.447±0.357(1)	.390±.007(3)	.942±.001(1)	.326±.006(1)	.924±.002(1)	.435±.004(2)	.938±.002(1)