Impact of SoLID data on quark transversity distributions

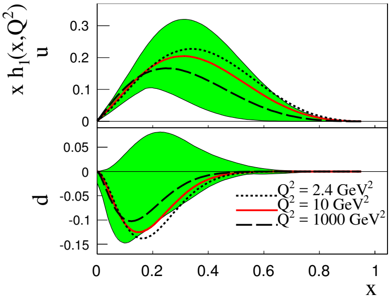
Nobuo Sato

In Collaboration with:
Z. Ye, K. Allada, T. Liu, A. Prokudin

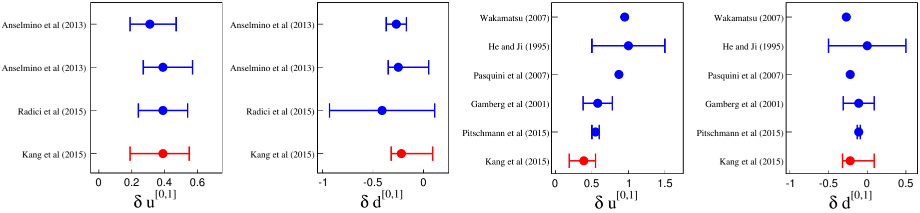
Jefferson Lab
EXPLORING THE NATURE OF MATTER

SoLID Collaboration Meeting, August, 2016

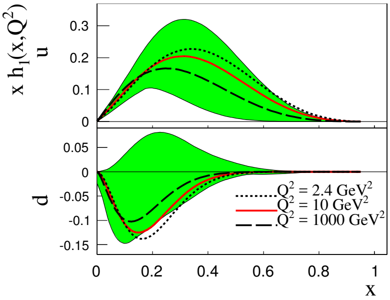
KPSY15 Analysis (PRD.93.014009)



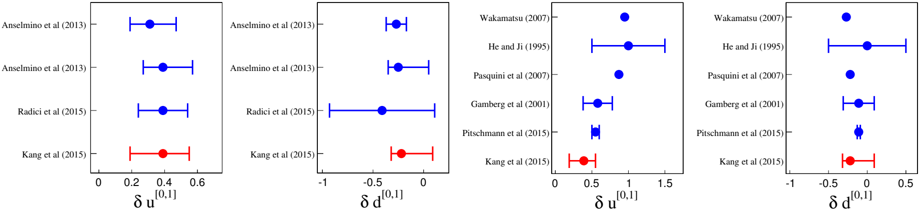
- Theory $\rightarrow$TMD factorization (No Y term).
- Observables: Collins Asymmetries from SIDIS ans SIA.
- Fitting: Maximun Likelihood (ML) $+$ CL based on $\Delta\chi^2$ envelope method.



KPSY15 Analysis (PRD.93.014009)



- Theory $\rightarrow$TMD factorization (No Y term).
- Observables: Collins Asymmetries from SIDIS ans SIA.
- Fitting: Maximun Likelihood (ML) $+$ CL based on $\Delta\chi^2$ envelope method.
- **...Impact of future SoLID data ?**



Theory of fitting

The goal is to estimate:

$$\mathrm{E}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; \mathcal{O}(\boldsymbol{a})$$

$$\mathrm{V}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; [\mathcal{O}(\boldsymbol{a}) - \mathrm{E}[\mathcal{O}]]^2$$

- $\mathcal{O}$=$h_u, \;\; h_d, \;\; \delta u, \;\; \delta d$
- $\sqrt{V[\mathcal{O}]} = 1\sigma$
- $\boldsymbol{a}$ $\rightarrow$ model parameters
- $\mathcal{P}(\boldsymbol{a}|data) \propto \mathcal{L}(data|\boldsymbol{a})\pi(\boldsymbol{a})$
- $\mathcal{L}(data|\boldsymbol{a}) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$
- $\chi^2(\boldsymbol{a}) = \sum_i \left(\frac{D_i - T_i(\boldsymbol{a})}{\delta D_i}\right)^2$
- $\pi(\boldsymbol{a})$ $\rightarrow$*priors.* i.e $\prod_i \theta(a_i - a_i^0)$

Theory of fitting

The goal is to estimate:

$$\mathrm{E}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; \mathcal{O}(\boldsymbol{a})$$

$$\mathrm{V}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; [\mathcal{O}(\boldsymbol{a}) - \mathrm{E}[\mathcal{O}]]^2$$

- $\mathcal{O}= h_u, \;\; h_d, \;\; \delta u, \;\; \delta d$
- $\sqrt{V[\mathcal{O}]} = 1\sigma$
- $\boldsymbol{a} \rightarrow$ model parameters
- $\mathcal{P}(\boldsymbol{a}|data) \propto \mathcal{L}(data|\boldsymbol{a})\pi(\boldsymbol{a})$
- $\mathcal{L}(data|\boldsymbol{a}) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$
- $\chi^2(\boldsymbol{a}) = \sum_i \left(\frac{D_i - T_i(\boldsymbol{a})}{\delta D_i}\right)^2$
- $\pi(\boldsymbol{a}) \rightarrow$ *priors*. i.e $\prod_i \theta(a_i - a_i^0)$
- **How to evaluate $E[\mathcal{O}], V[\mathcal{O}]$?**

Theory of fitting

The goal is to estimate:

$$\mathrm{E}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; \mathcal{O}(\boldsymbol{a})$$

$$\mathrm{V}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; [\mathcal{O}(\boldsymbol{a}) - \mathrm{E}[\mathcal{O}]]^2$$

Monte Carlo methods

- $\mathcal{P}(\boldsymbol{a}|data) \rightarrow \{\boldsymbol{a}_k\}$
- $\mathrm{E}[\mathcal{O}] = \frac{1}{N}\sum_k \mathcal{O}(\boldsymbol{a}_k)$
- $\mathrm{V}[\mathcal{O}] = \frac{1}{N}\sum_k [\mathcal{O}(\boldsymbol{a}_k) - \mathrm{E}[\mathcal{O}]]^2$

Maximun Likelyhood

- Maximize $\mathcal{P}(\boldsymbol{a}|data) \rightarrow \boldsymbol{a}_0$
- $\mathrm{E}[\mathcal{O}] \approx \mathcal{O}(\boldsymbol{a}_0)$
- $\mathrm{V}[\mathcal{O}] \approx$hessian, $\Delta\chi^2$ envelope,...

Theory of fitting

The goal is to estimate:

$$\mathrm{E}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; \mathcal{O}(\boldsymbol{a})$$

$$\mathrm{V}[\mathcal{O}] = \int d^n a \;\; \mathcal{P}(\boldsymbol{a}|data) \;\; [\mathcal{O}(\boldsymbol{a}) - \mathrm{E}[\mathcal{O}]]^2$$

Monte Carlo methods

- $\mathcal{P}(\boldsymbol{a}|data) \rightarrow \{\boldsymbol{a}_k\}$
- $\mathrm{E}[\mathcal{O}] = \frac{1}{N} \sum_k \mathcal{O}(\boldsymbol{a}_k)$
- $\mathrm{V}[\mathcal{O}] = \frac{1}{N} \sum_k [\mathcal{O}(\boldsymbol{a}_k) - \mathrm{E}[\mathcal{O}]]^2$

Maximun Likelyhood

- Maximize $\mathcal{P}(\boldsymbol{a}|data) \rightarrow \boldsymbol{a}_0$
- $\mathrm{E}[\mathcal{O}] \approx \mathcal{O}(\boldsymbol{a}_0)$
- $\mathrm{V}[\mathcal{O}] \approx$ hessian, $\Delta\chi^2$ envelope,...

The Hessian method

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$
- $\chi^2(\boldsymbol{a}) = \chi^2(\boldsymbol{a}_0) + \Delta\chi^2(\boldsymbol{a})$
- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\Delta\chi^2(\boldsymbol{a})\right)$

The Hessian method

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$

- $\chi^2(\boldsymbol{a}) = \chi^2(\boldsymbol{a}_0) + \Delta\chi^2(\boldsymbol{a})$

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\Delta\chi^2(\boldsymbol{a})\right)$

$$H_{ij} = \frac{1}{2}\frac{\partial\chi^2(\boldsymbol{a})}{\partial a_i \partial a_j}\bigg|_{\boldsymbol{a}=\boldsymbol{a_0}} = C^{-1}_{i,j}$$

$$H\hat{e}_k = w_k \hat{e}_k$$

$$\Delta\boldsymbol{a}^{(k)} = \boldsymbol{a}^{(k)} - \boldsymbol{a}_0 = t_k \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}$$

- $\Delta\chi^2(\boldsymbol{a}) \approx \Delta\boldsymbol{a}^T\, H\, \Delta\boldsymbol{a} = \sum_k \left(t_k \frac{\hat{\boldsymbol{e}}_k^T}{\sqrt{w_k}}\right) w_k \left(t_k \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right) = \sum_k t_k^2$

The Hessian method

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$
- $\chi^2(\boldsymbol{a}) = \chi^2(\boldsymbol{a}_0) + \Delta\chi^2(\boldsymbol{a})$
- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\Delta\chi^2(\boldsymbol{a})\right)$

$$H_{ij} = \frac{1}{2}\frac{\partial\chi^2(\boldsymbol{a})}{\partial a_i \partial a_j}\bigg|_{\boldsymbol{a}=\boldsymbol{a_0}} = C^{-1}_{i,j}$$

$$H\hat{e}_k = w_k \hat{e}_k$$

$$\Delta\boldsymbol{a}^{(k)} = \boldsymbol{a}^{(k)} - \boldsymbol{a}_0 = t_k \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}$$

- $\Delta\chi^2(\boldsymbol{a}) \approx \Delta\boldsymbol{a}^T H \Delta\boldsymbol{a} = \sum_k \left(t_k \frac{\hat{\boldsymbol{e}}_k^T}{\sqrt{w_k}}\right) w_k \left(t_k \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right) = \sum_k t_k^2$
- $\mathcal{P}(\boldsymbol{a}|data) \propto \prod_k \exp\left(-\frac{1}{2}t_k^2\right) \rightarrow$ $t_k = 1$ **gives** 1σ
- $\mathcal{P}(\boldsymbol{a}|data) \propto \prod_k \mathcal{P}_k(t_k|data) \rightarrow$ **factorization along eigen directions**

The Hessian method

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a})\right)$

- $\chi^2(\boldsymbol{a}) = \chi^2(\boldsymbol{a}_0) + \Delta\chi^2(\boldsymbol{a})$

- $\mathcal{P}(\boldsymbol{a}|data) \propto \exp\left(-\frac{1}{2}\Delta\chi^2(\boldsymbol{a})\right)$

$$H_{ij} = \frac{1}{2}\frac{\partial\chi^2(\boldsymbol{a})}{\partial a_i \partial a_j}\bigg|_{\boldsymbol{a}=\boldsymbol{a_0}} = C^{-1}_{i,j}$$

$$H\hat{e}_k = w_k\hat{e}_k$$

$$\Delta\boldsymbol{a}^{(k)} = \boldsymbol{a}^{(k)} - \boldsymbol{a}_0 = t_k\frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}$$

- $\Delta\chi^2(\boldsymbol{a}) \approx \Delta\boldsymbol{a}^T\, H\, \Delta\boldsymbol{a} = \sum_k \left(t_k\frac{\hat{\boldsymbol{e}}_k^T}{\sqrt{w_k}}\right) w_k \left(t_k\frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right) = \sum_k t_k^2$

- $\mathcal{P}(\boldsymbol{a}|data) \propto \prod_k \exp\left(-\frac{1}{2}t_k^2\right) \rightarrow$ $t_k = 1$ **gives** 1σ

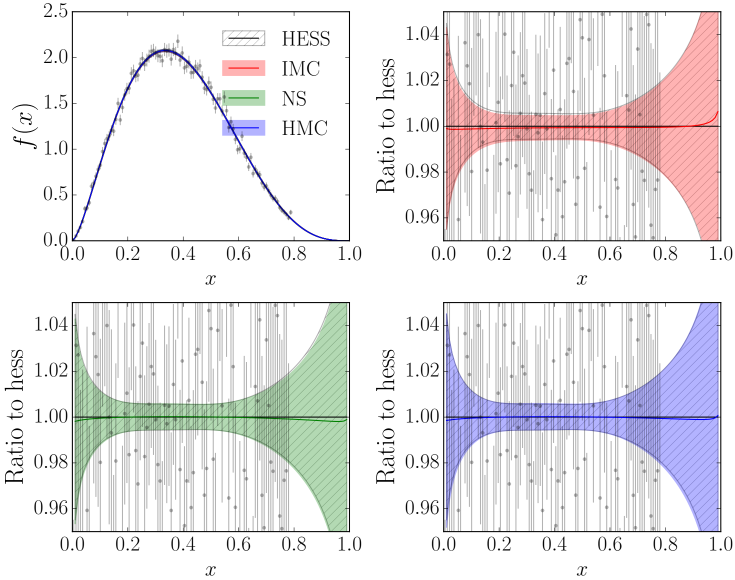
- $\mathcal{P}(\boldsymbol{a}|data) \propto \prod_k \mathcal{P}_k(t_k|data)$ $\rightarrow$**factorization along eigen directions**

- $V[\mathcal{O}] \approx \sum_k \delta\mathcal{O}_k^2 = \frac{1}{4}\sum_k\left[\mathcal{O}\left(\boldsymbol{a}_0 + \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right) - \mathcal{O}\left(\boldsymbol{a}_0 - \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right)\right]^2$

Simple 2D example

- $f(x) = x^{\alpha}(1-x)^{\beta}$. We set $\alpha = 1.5$ and $\beta = 3$.
- Generate pseudo data (rejection-sampling)
- Estimate $\mathrm{E}[f(x)]$ and $\mathrm{V}[f(x)]$ using:
  - ML + Hessian
  - Iterative Monte Carlo Method (IMC)
  - Hamiltonian Markov chain (HMC)
  - Nested sampling (NS)

Simple 2D example



Ambiguities/Confusion

The tolerance criterion

- ad-hoc inflation of errors via tolerance factor T
- $V[\mathcal{O}] \approx \sum_k \delta\mathcal{O}_k^2 = \frac{\mathrm{T}^2}{4} \sum_k \left[\mathcal{O}\left(\boldsymbol{a}_0 + \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right) - \mathcal{O}\left(\boldsymbol{a}_0 - \frac{\hat{\boldsymbol{e}}_k}{\sqrt{w_k}}\right)\right]^2$
- It can be shown that $\mathrm{T}^2 = \Delta\chi^2$
- In the example $\mathrm{T}^2 = \Delta\chi^2 = 1$ $\rightarrow$consistent with MC approaches

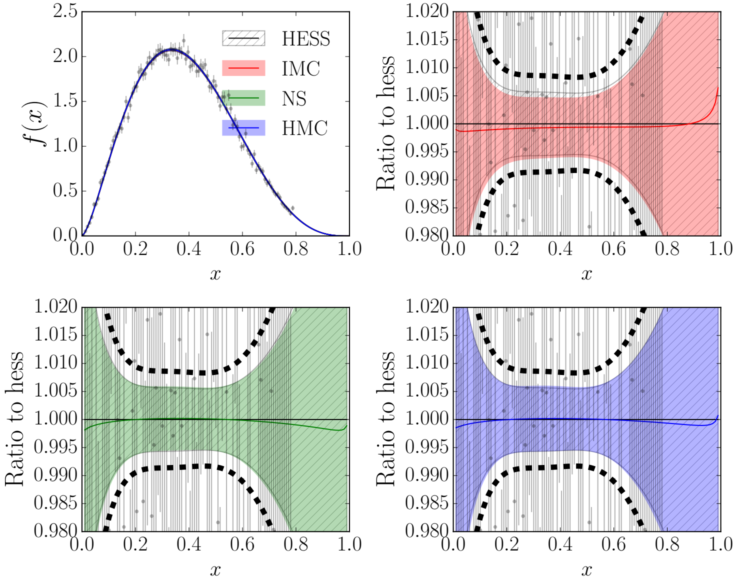
Claim by some groups (N =d.o.f)

$$68\%\ \mathrm{CL} = \int_0^{\Delta\chi^2} \frac{d\chi^2}{2\Gamma\left(\frac{N}{2}\right)} \left(\frac{\chi^2}{2}\right)^{\frac{N}{2}-1} \exp\left[-\frac{\chi^2}{2}\right]$$

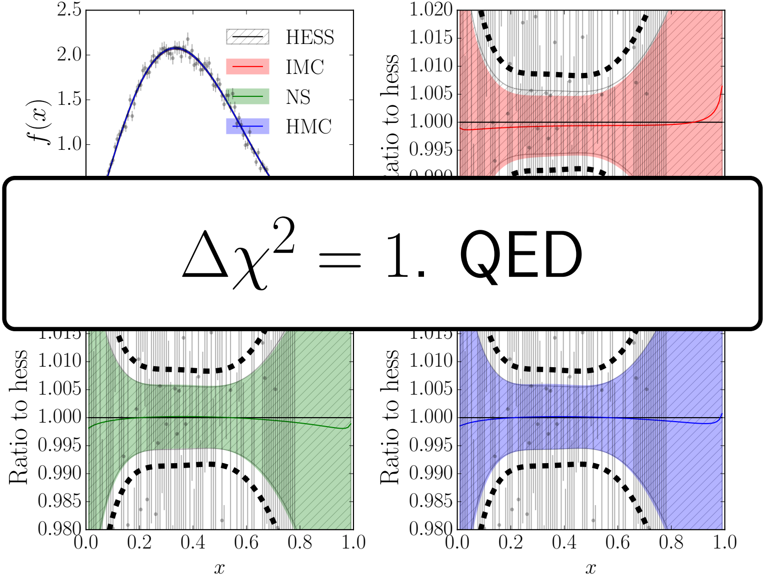
- $N = 1 \rightarrow \Delta\chi^2 = 1$
- $N = 2 \rightarrow \Delta\chi^2 = 2.3$
- $N = 3 \rightarrow \Delta\chi^2 = 3.53$

KPSY15 $\rightarrow \Delta\chi^2 = 29.7$ for 90%CL

Simple 2D example: $\Delta\chi^2 = 2.3$



Simple 2D example: $\Delta\chi^2 = 2.3$



$\Delta\chi^2 = 1.$ QED

A recipe to estimate impact of future data

How to estimate impact using projected uncertainties?

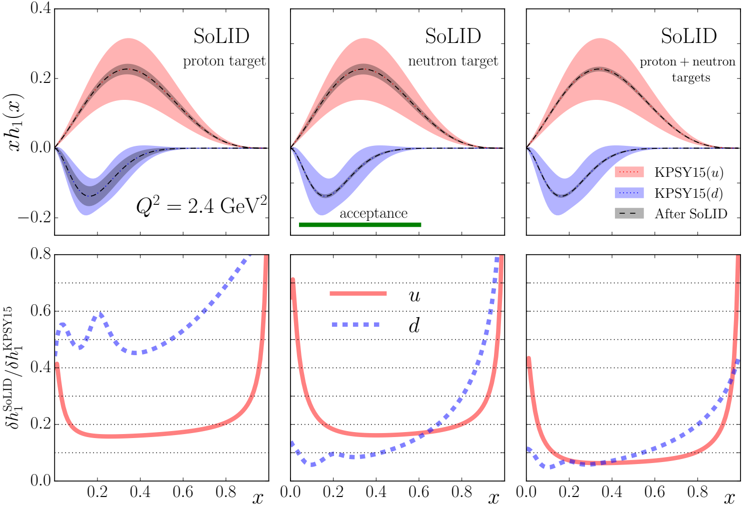
$$\chi^2(\boldsymbol{a}) = \sum_i \left(\frac{D_i^{\rm OLD} - T_i(\boldsymbol{a})}{\delta D_i^{\rm OLD}}\right)^2 + \sum_i \left(\frac{D_i^{\rm SoLID} - T_i(\boldsymbol{a})}{\delta D_i^{\rm SoLID}}\right)^2$$

$$H_{i,j} = \frac{1}{2}\frac{\partial \chi^2(\boldsymbol{a}, D)}{\partial a_i \partial a_j}\bigg|_{\boldsymbol{a}_0} = H_{i,j}^{\rm OLD} + H_{i,j}^{\rm SoLID} = C_{\rm OLD}^{-1} + H_{i,j}^{\rm SoLID}$$

Comments

- Only the Hessian of the new data set is needed.
- χ^2 is additive $\rightarrow$the Hessian can be partitioned (proton, neutron, kinematics,etc..).
- Different combinations of the Hessian can used to see what measurements are more relevant.

Results (See Talk by Tianbo)



To be consistent with KPSY15 $\rightarrow \Delta\chi^2 = 29.7$ for 90%CL

Delineating Current vs. Target fragmentation regions

In Collaboration with:
T. Rogers, O. Gonzalez

Current and target regions

Why do we care about the current region?

- Is the region where TMD fragmentation functions can be measured/ useful.

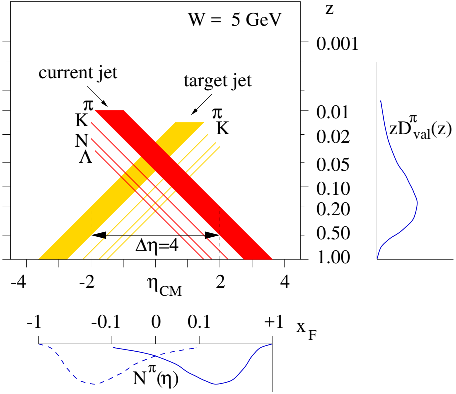
Experimental point of view

- Target region $\rightarrow$identifiable in the hadron rapidity distribution.
- Current region $\rightarrow$what is not in the target region.

Theory point of view

- Current region is a kinematic phase space where factorization theorems are applicable.

The criterion



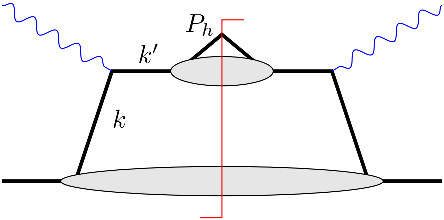
P.Mulders (hep-ph/0010199):

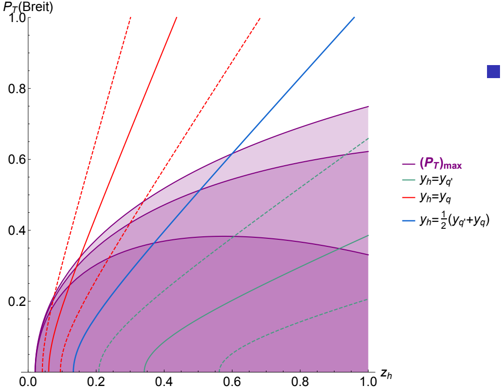
- $z = z_c = \frac{P_B^-}{q^-}$ in the current region
- $z = z_t = \frac{P_B^-}{(1-x_b)P^+}$ in the target region
- "Looking at the $\Delta\eta = 4$ difference one can estimate z-values above which current fragmentation dominates".

Critique

- $z = \frac{P\cdot P_b}{P\cdot q} \approx \frac{P_h^-}{q^-}$ ans **is never equal to** z_t
- $\Delta\eta$ seems to have no meaning.

New approach (Rogers, et al)





Conditions for factorization

- $P_h \cdot k = \mathcal{O}(Q^2)$
- $P_h \cdot k' = \mathcal{O}(m^2)$

Improved criteria

- Delineation of current region→consistent with factorization

Outlook/summary

- The hessian error analysis $\rightarrow$tested against MC methods.
- Clarification regarding $\Delta\chi^2 = 1$.
- A recipe to explore impact of future data sets on TMD distributions.
- Future SoLID data reduces the uncertainties on transversity (u, d) upto 90% for $x \in [0.05, 0.6]$
- Issues regarding the separation of current and target fragmentation.
- New criteria consistent with TMD factorization (preliminary).