Supplementary Material for Perception Preserving Projections

BMVC 2013 Submission # 269

Recall our objective function of PPP:

$$\min_{U^T U = I_r} \mathcal{L}(U) = \|\mathcal{P}(X) - \mathcal{P}(UU^T X)\|_F^2, \tag{1}$$

In this supplementary material, we briefly illustrate how we develop the gradient descent method on the Stiefel manifold defined by $U^T U = I$ [1]. The whole optimization is performed iteratively. In each iteration, we first find the gradient of the objective function defined in (1) in the tangent plane of the manifold at the solution point in the last iteration. Then a curve along the projected negative gradient is found and through a curvilinear search the solution for the next iteration is determined.

The gradient of the objective function in (1) at solution U^k is calculated as follows:

$$\nabla F(U^k) = (G{U^k}^T - U^k G^T)U^k, \tag{2}$$

where $G = (AU^k{U^k}^T B + BU^k{U^k}^T A - AB - BA)U^k$, $A = XX^T$, $B = P^T P$. Following [3], the curve along which search for the next solution can be defined by performing Cayley transformation [2] over the current solution:

$$Y(\tau) = (I + \frac{\tau}{2}W^k)^{-1}(I - \frac{\tau}{2}W^k)U^k,$$

where $W^k = G{U^k}^T - U^k G^T$ and G is defined as above. By applying the Sherman-Morrison-Woodbury formula [4], the above Cayley transformation can be simplified as:

$$Y(\tau) = X - \tau R(I + \frac{\tau}{2}V^T R)^{-1} V^T U^k, \tag{3}$$

where $R = [G, U^k]$ and $V = [U^k, -G]$. The curvilinear search along the above curve is performing traditional linear search until the Armijo-Wolfe conditions are satisfied. The details are presented in Algorithm 1.

Throughout the parameters are fixed as $\rho_1 = 1 \times 10^{-8}, \rho_2 = 1 \times 10^{-5}$ and τ is intialized as 1×10^{-3} to achieve the best performance.

References

[1] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. 2009.

Algorithm 1: Stiefel manifold gradient descent for PPP optimization.

Input : X, $\mathcal{P}$, parameter $0 < \rho_1 < \rho_2 < 1$.
Output: Reconstruction basis U.

repeat
  (S.1) Calculate the gradient according to (2).
  (S.2) Calculate the curve as in (3).
  (S.3) Initialize τ to a non-zero value.
  Calculate $F'(Y(\tau)) = \mathrm{tr}(G^T Y'(\tau))$, where $Y'(\tau) = -\left(I + \frac{\tau}{2}A\right)^{-1} A \left(\frac{U+Y(\tau)}{2}\right)$,
  $Y'(0) = -AX$, $A = GU^T - UG^T$.
  while $F(Y(\tau)) > F(Y(0)) + \rho_1 \tau F'(Y(0))$ *and* $F'(Y(\tau)) < \rho_2 F'(Y(0))$ **do**
    $\tau \leftarrow \frac{\tau}{2}$
  end
until *Convergence*;

[2] F Diele, L Lopez, and R Peluso. The cayley transform in the numerical solution of unitary differential systems. *Advances in computational mathematics*, 8(4):317–334, 1998.

[3] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, pages 1–38, 2010.

[4] Wikipedia. Sherman morrison formula, 2013.